

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

**THIS PAGE BLANK (USPTO)**



REC'D 29 NOV 1999

WIPO PCT

09/831426

FR 99 / 12738 4

# BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

COPIE OFFICIELLE

**PRIORITY  
DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le **16 AOUT 1999**

Pour le Directeur général de l'Institut  
national de la propriété industrielle  
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

## SIEGE

26 bis, rue de Saint Petersburg  
75800 PARIS Cédex 08  
Téléphone : 01 53 04 53 04  
Télécopie : 01 42 93 59 30

**THIS PAGE BLANK (USPTO)**

**REQUÊTE EN DÉLIVRANCE**

26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08  
Téléphone : (1) 42.94.52.52 Télécopie : (1) 42.93.59.30

Confirmation d'un dépôt par télécopie ☐

Cet imprimé est à remplir à l'encre noire en lettres capitales

Réservé à l'INPI

DATE DE REMISE DES PIÈCES

10 NOV 1998

N° D'ENREGISTREMENT NATIONAL

98 14146

DÉPARTEMENT DE DÉPÔT

DATE DE DÉPÔT

10 NOV. 1998

2 DEMANDE Nature du titre de propriété industrielle

☐ brevet d'invention ☐ demande divisionnaire

☐ certificat d'utilité ☐ transformation d'une demande de brevet européen

☐ demande initiale

☐ brevet d'invention

Établissement du rapport de recherche

☐ différé ☒ immédiat

Le demandeur, personne physique, requiert le paiement échelonné de la redevance

☐ oui ☒ non

Titre de l'invention (200 caractères maximum)

Gène humain htFIIIA et la protéine codée htFIIIA.

1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE  
À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE

Hoechst Marion Roussel  
Monsieur VIEILLEFOSSE Jean Claude  
102, Route de Noisy  
93235 ROMAINVILLE CEDEX

n° du pouvoir permanent PG06335 références du correspondant ML/2503 téléphone 0149915727

☐ certificat d'utilité n° date

3 DEMANDEUR (S) n° SIREN 5 5 2 0 8 1 4 7 3 code APE-NAF

Nom et prénoms (souligner le nom patronymique) ou dénomination

Hoechst Marion Roussel

Forme juridique

Société Anonyme à  
Directoire et Conseil  
de Surveillance

Nationalité (s) FRANCAISE

Adresse (s) complète (s)

1, Terrasse Bellini  
92800 PUTEAUX

Pays

FRANCE

En cas d'insuffisance de place, poursuivre sur papier libre ☐

4 INVENTEUR (S) Les inventeurs sont les demandeurs

☐ oui ☒ non Si la réponse est non, fournir une désignation séparée

5 RÉDUCTION DU TAUX DES REDEVANCES

☐ requise pour la 1ère fois ☐ requise antérieurement au dépôt : joindre copie de la décision d'admission

6 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE

pays d'origine

numéro

date de dépôt

nature de la demande

7 DIVISIONS antérieures à la présente demande n°

date

n°

date

8 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE

(nom et qualité du signataire - n° d'inscription)

Jean Claude VIEILLEFOSSE

SIGNATURE DU PRÉPOSÉ À LA RÉCEPTION

SIGNATURE APRES ENREGISTREMENT DE LA DEMANDE À L'INPI

# DOCUMENT COMPORTANT DES MODIFICATIONS

PAGE(S) DE LA DESCRIPTION OU DES REVENDECATIONS OU PLANCHE(S) DE DESSIN			R.M.*	DATE DE LA CORRESPONDANCE	TAMPON DATEUR DU CORRECTEUR
Modifiée(s)	Supprimée(s)	Ajoutée(s)			
16, 27, 28 + Dessins avec liste de références			X	21.05.99.	01 JUIN 1999 - 99

Un changement apporté à la rédaction des revendications d'origine, sauf si celui-ci découle des dispositions de l'article R.612-36 du code de la Propriété Intellectuelle, est signalé par la mention «R.M.» (revendications modifiées).

Gène humain htFIIIA et la protéine codée hTFIIIA

La présente invention concerne le gène codant pour le facteur de transcription humain nommé ci-après gène htFIIIA (ou htfc2) et la protéine codée hTFIIIA ainsi que l'utilisation de ce gène htFIIIA et celui de la protéine codée hTFIIIA dans le diagnostic et l'identification de certaines maladies liées au mécanisme de la transcription.

Nous appellerons ci-après tFIIIA (ou tfc2) le gène codant pour le facteur de transcription TFIIIA et htFIIIA le gène codant pour le facteur de transcription humain hTFIIIA. Le gène humain htFIIIA code donc pour la protéine correspondante hTFIIIA.

Nous utiliserons également ci-après les abréviations suivantes : AA pour acides aminés, AN pour acides nucléiques, pb pour paires de bases, ADN pour acide désoxyribonucléique, ADNc pour ADN complémentaire, ARN pour acide ribonucléique et RNase pour ribonucléase et C pour désoxycytidine.

On utilisera également le terme screening qui désigne une technique de criblage spécifique et le terme primer qui désigne un oligonécléotide utilisé en amorce. Le gène tFIIIA et la protéine correspondante TFIIIA seraient impliqués dans la régulation du mécanisme biologique de la transcription comme indiqué ci-après.

Depuis que la protéine TFIIIA a été purifiée comme facteur de transcription pour la première fois en 1980 à partir d'ovocytes de Xénope [Segall et al, J. Biol. Chem., 255, 11986-11991 (1980)], des travaux ont été menés in vivo et in vitro dans le Xénope pour étudier le mécanisme de contrôle de la transcription exercé par TFIIIA. On a ainsi montré que TFIIIA de Xénope est nécessaire pour l'initiation de la transcription du gène ARN 5S [Sakonji et al, Cell 19, 13-25 (1980)] et se lie à une région de contrôle interne du gène ARN 5S [Bogenhagen et al, Cell, 19, 27-35 (1980)].

La séquence en nucléotides de l'ADNc de tFIIIA de Xénope et la séquence correspondante d'acides aminés ont déjà été publiées [Ginberg et al, Cell 39, 479-489 (1984)]. On peut noter que ce gène code pour une structure en 9 doigts de

zinc, un doigt de zinc correspondant à la répétition du motif CYS2 HIS2 (C2H2). Cette structure en doigt de zinc est considérée comme un domaine essentiel pour un groupe de protéines qui se lient à l'ADN (DNA binding proteins) [Miller et al, Embo J., 4, 1607-1614 (1985)].

On connaît ainsi dans l'être humain des facteurs de transcription se liant à l'ADN qui possèdent également cette structure en doigt de zinc tels que par exemple XT1 du gène de la tumeur humaine de Wilms, [Gessier et al, Nature, 343, 774-778 (1990)], le répresseur humain de transcription YY1 [SHI et al, Cell, 67, 377-388 (1991)], la protéine MAZ associée au gène humain MYC [Bossone et al, Proc. Natl. Acad. Sci., USA, 89, 7452-7456 (1992)] ou encore spl [Kuwahara et al, J.Biol. Chem., 265, 8627-8631 (1990)].

Des travaux ont été effectués afin d'isoler le gène humain htflIIIA, mais aucun jusqu'ici n'a abouti à mettre en évidence la véritable séquence du gène htflIIIA.

On peut citer ainsi d'une part les travaux décrits dans la demande européenne EP 0704526 (Fujisawa et al), et repris dans l'article : Arakawa et al (1995), Cytogenet Cell Genet 70,235-238, qui ont abouti à une séquence que nous appellerons htflIIIA de Arakawa et d'autre part les travaux décrits dans l'article : DREW et al (1995), Gene 159,215-218, qui ont abouti à une séquence que nous appellerons htflIIIA de DREW. Ces séquences htflIIIA de DREW et de ARAKAWA sont représentées respectivement aux figures 4 et 5 ci-après.

Les documents indiqués ci-dessus décrivent donc chacun une séquence du gène htflIIIA mais ces deux séquences diffèrent l'une de l'autre par quelques nucléotides et diffèrent du gène htflIIIA de la présente demande comme indiqué ci-après.

La présente invention a permis d'isoler le gène codant pour le facteur de transcription humain hTfIIIA.

La présente invention a également permis de révéler la séquence d'acides nucléiques du gène htflIIIA et également la séquence d'acides aminés de la protéine hTfIIIA codée par ce gène.

La présente invention a ainsi pour objet la séquence d'ADN du gène htflIIIA codant pour une protéine ayant la



fonction biologique du facteur de transcription humain hTFIIIA.

La présente invention a précisément pour objet la séquence d'ADN du gène htfIIIA du facteur de transcription humain hTFIIIA telle que définie ci-dessus codant pour la  
5 séquence d'acides aminés SEQ ID N°2.

Une telle séquence SEQ ID n°2 de la présente invention comprend donc 365 acides aminés.

La présente invention a aussi pour objet la séquence  
10 d'ADN du gène htfIIIA comme défini ci-dessus correspondant à la séquence de nucléotides SEQ ID N°3.

La séquence SEQ ID N°3 comprend donc 1273 nucléotides.

La présente invention a notamment pour objet la séquence d'ADN du gène htfIIIA comme défini ci-dessus correspondant à  
15 la séquence de nucléotides SEQ ID N°4. La séquence SEQ ID N°4 comprend donc 1213 nucléotides.

La séquence SEQ ID N°1 représente sur la ligne supérieure la séquence de nucléotides du gène htfFIIIA selon la présente invention soit SEQ ID N°3, et sur la ligne inférieure, la  
20 séquence correspondante en acides aminés (AA) de cette séquence de nucléotides soit SEQ ID N°2.

Les figures 1 et 2 ci-après représentent sur la ligne supérieure la séquence d'AA codée par htfIIIA de la présente invention SEQ ID N°2 et sur la ligne inférieure, les  
25 séquences d'AA codées respectivement par les gènes htfIIIA de DREW, à la figure 1, et de ARAKAWA, à la figure 2, ces séquences de DREW et ARAKAWA étant telles que publiées dans les documents référencés ci-dessus.

La figure 3 ci-après représente la comparaison des  
30 séquences d'AA codées respectivement par les gènes htfIIIA de DREW et de ARAKAWA avec sur la ligne supérieure, la séquence d'AA codée par htfIIIA de ARAKAWA et sur la ligne inférieure, la séquence d'AA codée par htfIIIA de DREW.

La figure 2 montre donc que la séquence d'AA  
35 correspondante de htfIIIA selon la présente invention comporte des différences avec la séquence d'AA publiée dans l'article de ARAKAWA ou dans EP 0704 526, notamment aux positions respectives correspondantes 105 et 163, 156 et 214,

320 à 329 et 378 à 387, ces positions étant données par rapport à la numérotation indiquée sur la figure 2.

Cette figure 2 montre également que la séquence d'AA codée par htFIIIA de la présente invention débute en position 59 de la séquence en AA de htFIIIA de ARAKAWA.

La figure 3 montre que les séquences d'AA codées par htFIIIA de ARAKAWA et de DREW comportent des différences aux positions respectives correspondantes 214 et 154, 378-387 et 318-327, ces positions étant données par rapport à la numérotation indiquée sur la figure 3.

La figure 5 montre que la séquence htFIIIA de ARAKAWA code pour une protéine dont la séquence en acides aminés, indiquée dans EP 0704 526, commence par l'AA méthionine spécifié par le codon ATG qui se trouve en position 20-22 et la traduction s'arrête à un codon TAA. Si l'on compare la séquence de nucléotides de htFIIIA selon la présente invention SEQ ID N°3 avec la séquence de nucléotides de EP 0704 526 soit htFIIIA de ARAKAWA représentée à la figure 5 (séquence p11-12-13 de EP 0704 526), on constate qu'il manque un nucléotide C à la position 127 de la séquence de EP 0704 526. Ce nucléotide C supplémentaire a pour conséquence un décalage dans la traduction en acides aminés de cette séquence de nucléotides : en effet, l'ATG qui se trouve à la position 20-22 de la séquence de ARAKAWA représentée à la figure 5 et qui est considéré comme codon initiateur de la synthèse protéique par ARAKAWA n'est donc plus, du fait de ce décalage, dans la même phase de lecture. En tenant compte de ce nucléotide supplémentaire C, la traduction en AA fait apparaître un codon de terminaison TGA en position 57-59 de la séquence de ARAKAWA représentée à la figure 5. Par conséquent, le codon initiateur de la synthèse protéique selon la présente invention est localisée en aval de ce codon de terminaison. Des expériences de traduction in vitro de SEQ ID N°4 et des tests d'expression dans des cellules de mammifères telles que des cellules Cos ont permis d'identifier le codon d'initiation de la synthèse protéique de htFIIIA selon la présente invention.

Ce codon initiateur de la synthèse protéique de htFIIIA

selon la présente invention est le codon CTG en position 176-178 de SEQ ID N°3 (qui correspondrait à la position 194-196 de la séquence de ARAKAWA représentée à la figure 5).

La partie codante du gène htFIIIA de la présente invention commence donc par ce codon CTG qui se trouve à la position 176-178 de SEQ ID N°3 qui devrait correspondre à l'AA Leucine et qui correspond en fait à l'AA Méthionine puisque ce codon est reconnu comme codon initiateur (ref : David S. Peabody The Journal of Biological Chemistry, vol. 264, n°9, pp. 5031-5035, 1989).

En conséquence, la protéine htFIIIA de ARAKAWA est plus longue comme le montre la figure 2, que la protéine htFIIIA de la présente invention.

Par ailleurs, si l'on compare la protéine htFIIIA de la présente invention et la protéine htFIIIA de DREW, (comparaison représentée à la figure 1), on constate que l'acide aminé thréonine en position 105 de la protéine htFIIIA de la présente invention correspond à un résidu asparagine en position 103 dans la séquence htFIIIA de DREW et que les deux premiers AA, M et D de la protéine htFIIIA de la présente invention n'ont pas été déterminés pour la protéine htFIIIA de DREW. L'absence des codons spécifiant ces AA et notamment l'absence du codon initiateur de la synthèse protéique, ne permet pas l'expression de cette protéine. La séquence htFIIIA de DREW représentée à la figure 4 est donc incomplète, comme le reconnaissent les auteurs de la publication référencée ci-dessus (DREW et al à la page 216 lignes 39-41). On peut noter d'ailleurs que les auteurs de cet article pensent également que le codon initiateur de la séquence htFIIIA de DREW devrait correspondre à une méthionine codée par ATG comme dans la séquence de ARAKAWA.

Le gène htFIIIA selon la présente invention est donc différent des gènes htFIIIA de DREW et ARAKAWA (EP 0704526) et code pour une protéine htFIIIA dont la séquence en AA est différente de celle des protéines htFIIIA de DREW et ARAKAWA.

La présente invention a ainsi particulièrement pour objet la séquence d'ADN du gène htFIIIA comme défini ci-dessus contenant la séquence de nucléotides SEQ ID N°3.

La présente invention a plus particulièrement pour objet la séquence d'ADN comme défini ci-dessus ayant la séquence commençant au nucléotide 176 et se terminant au nucléotide 1270 de SEQ ID N°3.

5 Une telle séquence de la présente invention commence donc à un codon CTG et comprend donc 1095 nucléotides.

La présente invention a également pour objet la séquence d'ADN codant pour le facteur de transcription humain hTFIIIA comme défini ci-dessus ainsi que les séquences d'ADN qui  
10 hybrident avec celle-ci et/ou présentent des homologies significatives avec cette séquence ou des fragments de celle-ci et codant pour des protéines ayant la même fonction. Par séquences qui hybrident, on inclut les séquences d'ADN qui hybrident avec l'une des séquences d'ADN ci-dessus sous  
15 des conditions standard de stringence élevée, moyenne ou basse. Par protéines ayant la même fonction, on inclut les polypeptides ayant la même fonction de facteur de transcription. Les conditions de stringence sont celles réalisées dans les conditions connues de l'homme du métier telles que celles  
20 décrites par Sambrook et al (1989) Molecular cloning, Cold Spring Harbor Laboratory Press, 1989. De telles conditions de stringence sont par exemple une hybridation à 65°C, pendant 18 heures dans une solution 5 x SSPE ; 10 x Denhardt's ; 100µg/ml ADNss ; 1 % SDS suivie de 3 lavages pendant 5 minutes avec 2 x SSC ; 0,05 % SDS, puis 3 lavages pendant 15  
25 minutes à 65°C dans 1 x SSC ; 0,1 % SDS. Les conditions de forte stringence comprennent par exemple une hybridation à 65°C, pendant 18 heures dans une solution 5 x SSPE ; 10 x Denhardt's ; 100µg/ml ADNss ; 1 % SDS suivie de 2 lavages pendant 20 minutes avec une solution 2 x SSC ; 0,05 % SDS à  
30 65°C suivis d'un dernier lavage pendant 45 minutes dans une solution 0,1 x SSC ; 0,1 % SDS à 65°C. Les conditions de stringence moyenne comprennent par exemple un dernier lavage pendant 20 minutes dans une solution 0,2 x SSC, 0,1 % SDS à  
35 65°C.

Par séquences qui présentent des homologies significatives, on inclut les séquences ayant une identité de séquence nucléotidique d'au moins 50 % avec l'une des séquences d'ADN

ci-dessus et qui codent pour une protéine ayant la même fonction de facteur de transcription.

La présente invention a également pour objet la séquence d'ADN comme défini ci-dessus comprenant des modifications  
5 introduites par suppression, insertion et/ou substitution d'au moins un nucléotide codant pour une protéine ayant la même activité biologique que le facteur de transcription humain hTFIIIA.

La présente invention a notamment pour objet la séquence  
10 d'ADN telle que définie ci-dessus ainsi que les séquences d'ADN similaires qui ont une homologie de séquence nucléotidique d'au moins 50 % ou au moins 60 % et de préférence au moins 70 % avec ladite séquence d'ADN.

La présente invention a ainsi également pour objet  
15 la séquence d'ADN telle que définie ci-dessus ainsi que les séquences d'ADN qui codent pour une protéine dont la séquence en AA a une homologie d'au moins 40 % et notamment de 45 % ou d'au moins 50 %, plutôt au moins 60 % et de préférence au moins 70 % avec la séquence en AA codée par ladite séquence  
20 d'ADN.

Le gène de la présente invention est représenté comme une séquence ADN simple brin mais il est entendu que la présente invention inclut la séquence ADN complémentaire de cette séquence ADN simple brin et inclut également la  
25 séquence ADN dite double brin constituée de ces deux séquences ADN complémentaires d'une de l'autre.

La séquence ADN de la présente invention est un exemple de combinaison de codons codant pour les acides aminés correspondant à la séquence d'acides aminés SEQ ID N°2 mais  
30 il est entendu également que la présente invention inclut toute autre combinaison arbitraire de codons codant pour cette même séquence d'acides aminés SEQ ID N°2.

La séquence d'ADN telle que définie ci-dessus ou cette séquence d'ADN modifiée comme indiquée ci-dessus, peut être  
35 préparée selon les techniques connues de l'homme du métier et notamment celles décrites dans l'ouvrage de Sambrook, J. Fritsh, E. F. & Maniatis, T. (1989) intitulé : « Molecular cloning : a laboratory manual », Laboratory, Cold Spring

Harbor NY. Notamment la séquence d'ADN ci-dessus peut être une séquence d'ADNc obtenue par identification des parties 3' et 5' de la séquence codante, puis amplification de ces parties à l'aide d'une ADN polymérase telles que la pfu polymérase ou d'autres ADN polymérases. L'introduction, dans la séquence des oligonucléotides utilisés pour la PCR, de sites de restriction tels que Hind III ou SmaI permet le clonage de ces fragments dans des vecteurs adéquats et ensuite la reconstitution de la séquence complète recherchée.

Une description détaillée des conditions opératoires dans lesquelles a été réalisée la présente invention est donnée ci-après.

L'invention a tout particulièrement pour objet le polypeptide ayant la fonction de facteur de transcription humain hTFIIIA et ayant la séquence d'acides aminés SEQ ID N°2 codée par la séquence d'ADN comme défini ci-dessus et les analogues de ce polypeptide.

Par analogues, on entend les polypeptides dont la séquence d'acides aminés a été modifiée par substitution, suppression ou addition d'un ou plusieurs acides aminés mais qui conservent la même fonction biologique. De tels polypeptides analogues peuvent être produits spontanément ou peuvent être produits par modification post-transcriptionnelle ou encore par modification de la séquence ADN de la présente invention comme indiqué ci-dessus, en utilisant les techniques connues de l'homme du métier : parmi ces techniques, on peut citer notamment la technique de mutagénèse dirigée (Kramer, W., et al., Nucl. Acids Res., 12, 9441 (1984) ; Kramer, W. and Fritz, H.J. , Methods in Enzymology, 154, 350 (1987) ; Zoller, M.J. and Smith, M. Methods in Enzymology, 100, 468 (1983)).

La synthèse d'ADN modifiés peut également être faite en utilisant des techniques de synthèse chimique bien connues telles que par exemple la méthode au phosphotriester [Letsinger, R.L and Ogilvie, K.K., K. Am. CHEM. Soc., 91, 3350 (1969) ; Merrifield, R.B., Sciences, 150, 178 (1968)] ou la méthode à la phosphoamidite [Beaucage, S.L and Caruthers, M .H., Tetrahedron Lett., 22, 1859 (1981) ; McBRIDE, L.J. and

Caruthers, M.H. Tetrahedron Lett., 24 245 (1983)] ou encore par la combinaison de ces méthodes.

Les polypeptides de la présente invention peuvent donc être préparés par les techniques connues de l'homme du  
5 métier, notamment partiellement par synthèse chimique ou encore par la synthèse de l'ADNc par expression dans une cellule hôte procaryote ou eucaryote comme indiqué ci-après.

La présente invention a particulièrement pour objet le procédé de préparation de la protéine recombinante hTFIIIA  
10 ayant la séquence d'acides aminés SEQ ID N°2. Ce procédé inclut l'expression de la séquence d'ADN comme défini ci-dessus dans un hôte approprié puis l'isolement et la purification de ladite protéine recombinante.

Pour produire le polypeptide de la présente invention,  
15 on peut notamment utiliser les techniques de l'ADN recombinant en utilisant les méthodes de génie génétique et de culture cellulaire connues de l'homme du métier. On peut ainsi procéder par les étapes suivantes : d'abord préparation du gène approprié, puis incorporation de ce gène dans un  
20 vecteur, transfert du vecteur porteur du gène dans une cellule hôte appropriée, expression du gène et enfin, purification de la protéine codée par ce gène.

Les séquences d'ADN selon la présente invention et notamment SEQ ID N°3 ou SEQ ID N°4 peuvent être préparées selon les  
25 techniques connues de l'homme du métier, notamment par synthèse chimique, par criblage d'une banque génomique ou d'une banque d'ADNc à l'aide de sondes d'oligonucléotides de synthèse par les techniques connues d'hybridation ou encore par reverse transcriptase à partir d'ARN messager (ARNm).

30 L'avantage de la technique comprenant d'abord l'isolement d'ARNm par extraction des ARN totaux puis la synthèse d'ADNc à partir de ces ARNm par reverse transcriptase réside notamment dans le fait que l'ARNm ne contient pas les introns alors que ces séquences non codantes sont présentes dans  
35 l'ADN génomique.

On peut donc notamment procéder comme suit.

On extrait d'abord les ARN totaux provenant d'une lignée cellulaire telle que par exemple la lignée cellulaire Raji

(RNA Plus, BIOPROBE) et à partir de ces ARN, on procède ensuite à la synthèse de l'ADNc recherché en utilisant notamment un kit tel que ARN PCR kit (Perkin Elmer).

On peut noter que dans le cadre de la présente invention, deux oligonucléotides localisés aux extrémités de la séquence codante htfIIIA publiée par ARAKAWA (figure 5) ont été synthétisés soit OLT5 et OLT3 définis comme suit :

5 - OLT5 : 5' CGGGGTACCAAAA ATG CGC AGC AGC GGC GCC GAC 3' soit SEQ ID N°5 et

10 - OLT3 : 5' CGGTCTAGA TTA GCC AAG GGT AAG TAC TGC 3' soit SEQ ID N°9

mais ces deux oligo-nucléotides n'ont pas permis d'obtenir un produit d'amplification par PCR.

Ainsi dans le cadre de la présente invention, la séquence codante de htfIIIA a été isolée en deux étapes : d'abord identification de la partie 3' puis identification de la partie 5'.

Après identification des parties 3' et 5', un site de restriction HindIII localisé sur chacun de ces fragments a ensuite permis de reconstituer la séquence complète recherchée comme indiqué ci-après dans la partie expérimentale. On a donc procédé comme suit :

La partie 3' a été amplifiée à l'aide de la pfu polymérase (STRATAGENE) en utilisant comme amorce les oligonucléotides

25 OLT5.2 et TFIIIA 3'SmaI soit :

- OLT5.2 : 5'TCCTTCCCTGACTGCAGCGCC 3' soit SEQ ID N°6 et

- TFIIIA3'SmaI : 5'CCT CCC GGG GCC AAG GGT AAG TAC TGC AAC 3'soit SEQ ID N°10

Les amorces d'amplification ou primers sont choisies en fonction de la partie à amplifier selon des critères usuels pour l'homme du métier.

Les amorces utilisées dans la présente invention ont été choisies dans la séquence htfIIIA de ARAKAWA représentée à la figure 5.

35 Les séquences SEQ ID N°6, SEQ ID N°7 et SEQ ID N°8 sont situées respectivement aux positions 320-340 (5'→3'), 361-380 (séquence réverse et complémentaire) et 391-410 (séquence



réverse et complémentaire) de cette séquence htFIIIA de ARAKAWA.

Les séquences SEQ ID N°5, SEQ ID N°9 et SEQ ID N°10 sont situées respectivement aux positions 20-40 (5'→3'), 1271

5 1291 (séquence réverse et complémentaire) et 1268-1288 (séquence réverse et complémentaire) de cette séquence htFIIIA de ARAKAWA.

On peut noter que dans les séquences SEQ ID N°5, SEQ ID N°9 et SEQ ID N°10 contiennent les séquences correspondant aux sites d'enzymes de restriction soit respectivement KpnI, XbaI  
10 et SmaI.

L'oligonucléotide TFIIIA 3' SmaI introduit un site de restriction SmaI en aval de la séquence codante. Ce site permet si nécessaire et si désiré la fusion de la séquence  
15 codant pour hTFIIIA avec une séquence codant pour un peptide épitope de l'hémagglutinine désigné « TAG HA ». L'expression de la séquence codant pour TFIIIA peut donc être associée à celle de la séquence codant pour TAG HA qui peut être détecté en analyse par Western blot, si le gène de fusion est  
20 exprimé.

Pour l'identification de la partie 5', cette région a été isolée par la technique de PCR ancrée 5' (5 race System, GIBCO BRL ; pfu polymerase, STRATAGENE). Deux PCR successives ont été effectuées en utilisant comme amorce les oligonu-  
25 cléotides suivants : UAP et TFIIIAPCR5' pour la première PCR et UAP et TFIIIA SEQ2 pour la seconde PCR. UAP est un oligonucléotide fourni dans le kit.

Ces oligonucléotides ont les séquences suivantes :

- TFIIIAPCR5' : 5' CACAAACAAATGGTCTCC 3' soit SEQ ID N°8
- 30 - TFIIIA SEQ2 : 5' TGCACAGGTGCGCGTCAAGC 3' soit SEQ ID N°7.

Les produits de ces PCR soit les fragments 5' et 3' amplifiés sont ensuite purifiés sur gel d'agarose et clonés en utilisant le TA cloning kit (INVITROGEN). On procède alors au séquençage : l'ADN plasmidique de plusieurs clones  
35 indépendants est préparé (QIAGEN Plasmids KIT) et les fragments correspondant à la séquence codante de hTFIIIA sont séquencés sur les deux brins (Séquenceur ABI 377XL, PERKIN ELMER).

On peut donc procéder comme suit et selon les techniques usuelles de clonage connues de l'homme du métier et notamment on procède au clonage par insertion de chaque fragment dans un plasmide fourni avec le kit commercial (TA cloning Kit  
5 Invitrogen) puis transformation d'une souche bactérienne par le plasmide ainsi obtenu. On peut utiliser notamment la souche E. coli XL1 Blue.

Les clones sont ensuite cultivés pour extraire l'ADN plasmidique selon les techniques classiques de l'homme du  
10 métier référencées ci-dessus (Sambrook, Fritsh et Maniatis).

On procède au séquençage de l'ADN du fragment amplifié contenu dans l'ADN plasmidique.

La compilation des données de la séquence ainsi obtenues fait apparaître qu'en 3', l'essentiel de la séquence isolée  
15 correspond à la séquence htflIIIA de Drew et al.

En 5', la séquence la plus longue débute en position 80 de la séquence htflIIIA de Arakawa et al., représentée à la figure 2F, et fait apparaître l'insertion d'un nucléotide C en position 127 par rapport à cette séquence. Si l'on peut  
20 supposer que la synthèse de l'ADNc dans l'application de la technique décrite ci-dessus n'a pas été complète, l'insertion d'un nucléotide crée cependant un problème majeur. En effet, l'addition d'un nucléotide dans la séquence codante crée un décalage de la phase de lecture. Afin de vérifier la présence  
25 de ce nucléotide dans le gène htflIIIA, de l'ADN génomique humain a été analysé par PCR. Cet ADN a été soumis à une réaction PCR à l'aide de la pfu polymérase (STRATAGENE) ou de la Taq polymérase (Perkin Elmer) et en utilisant comme amorce les oligonucléotides OLT5 et TFlIIIA SEQ2 nommés respecti-  
30 vement SEQ ID N°5 et SEQ ID N°7. Les deux produits PCR ont été clonés (TA cloning Kit) puis séquencés.

L'analyse des données de séquence confirme pour ces deux amplifications la présence de ce nucléotide C supplémentaire par rapport à la séquence htflIIIA de ARAKAWA. L'ATG décrit  
35 initialement comme codon initiateur de la synthèse protéique pour hTFlIIIA de ARAKAWA ne peut donc plus être considéré comme tel.

On procède ensuite à l'assemblage des séquences 5' et 3'

et obtient un plasmide unique contenant la séquence recherchée de hTFIIIA de la présente invention. On reconstitue la séquence codante hTFIIIA complète de la façon suivante. On digère un clone issu de l'amplification de l'ADN génomique à l'aide des enzymes de restriction EcoRI et HindIII, et on obtient après purification, un fragment de 350 pb environ. On digère par ailleurs un clone issu de l'amplification de la partie 3' à l'aide des enzymes de restriction HindIII et SmaI et on obtient après purification, un fragment de 930 pb environ.

On procède ensuite à la ligation de ces fragments dans le plasmide pYX223 (vecteur d'expression pour la levure - RSD) préalablement digéré par EcoRI et SmaI.

Un exposé détaillé des conditions dans lesquelles peuvent être menées les opérations indiquées ci-dessus est donné ci-après dans la partie expérimentale. On obtient ainsi un plasmide dans lequel est inséré le gène de la présente invention et on obtient ainsi également ce plasmide introduit dans une cellule hôte en opérant selon les techniques usuelles connues de l'homme du métier.

Le polypeptide de la présente invention peut être obtenu par expression dans une cellule hôte contenant la séquence d'ADN codant pour le polypeptide de l'invention précédée d'une séquence promotrice convenable. La cellule hôte peut être une cellule procaryote, par exemple E. coli ou une cellule eucaryote telle que les levures comme par exemple les ascomycètes parmi lesquels les Saccharomyces cerevisiae ou encore des cellules mammifères comme les cellules Cos.

La présente invention a particulièrement pour objet un vecteur d'expression contenant une séquence d'ADN comme défini ci-dessus.

Ainsi un tel vecteur d'expression selon la présente invention contient une séquence d'ADN qui peut être la séquence de nucléotides SEQ ID N°3 ou la séquence commençant au nucléotide 176 et se terminant au nucléotide 1270 de SEQ ID N°3.

Un tel vecteur d'expression selon la présente invention peut également contenir les séquences d'ADN qui hybrident

avec les séquences définies ci-dessus et/ou présentent des homologies significatives avec ces séquences ou des fragments de celles-ci.

5 Un tel vecteur d'expression selon la présente invention peut également contenir les séquences d'ADN qui comprennent des modifications introduites par suppression, insertion et/ou substitution d'au moins un nucléotide codant pour une protéine ayant la même activité biologique que le facteur de transcription humain hTFIIIA.

10 Les vecteurs d'expression sont des vecteurs permettant l'expression de la protéine sous le contrôle d'un promoteur convenable. Un tel vecteur peut être un plasmide, un cosmide ou un ADN viral. Pour les cellules procaryotes, le promoteur peut être par exemple le promoteur lac, le promoteur trp, le  
15 promoteur tac, le promoteur  $\beta$ -lactamase ou le promoteur PL. Pour les cellules de levure, le promoteur peut être par exemple le promoteur PGK ou le promoteur GAL. Pour les cellules de mammifères, le promoteur peut être par exemple le promoteur SV40 ou les promoteurs de l'adénovirus. Des  
20 vecteurs type Baculovirus peuvent être aussi utilisés pour l'expression dans des cellules d'insectes.

Les cellules hôtes sont par exemple des cellules procaryotes ou des cellules eucaryotes. Les cellules procaryotes sont par exemple E. coli, Bacillus ou Streptomyces. Les  
25 cellules hôtes eucaryotes comprennent des levures ainsi que des cellules d'organismes supérieurs, par exemple des cellules de mammifères ou des cellules d'insectes. Les cellules de mammifères sont par exemple des fibroblastes tels que des cellules CHO ou BHK de hamster et des cellules Cos de singe.  
30 Les cellules d'insectes sont par exemple des cellules SF9.

La présente invention concerne donc un procédé qui comprend l'expression de la protéine hTFIIIA dans une cellule hôte transformée par un ADN codant pour la séquence polypeptidique correspondant à la séquence SEQ ID N°2.

35 Pour la réalisation de la présente invention, les vecteurs utilisés peuvent être par exemple pGEX ou pBAD et la cellule hôte peut être E. coli ou par exemple le vecteur pYX223 et la cellule hôte peut être également S. cerevisiae.

La présente invention a notamment pour objet une cellule hôte transformée avec un vecteur tel que défini ci-dessus contenant le gène htfIIIA selon la présente invention.

La présente invention a très précisément pour objet le  
5 plasmide déposé à la CNCM sous le numéro I-2071.

Il s'agit ainsi de la souche XL1-Blue/pBShtfc2LHA renfermant le gène htfIIIA selon la présente invention.

Les conditions opératoires dans lesquelles a été réalisée la présente invention sont décrites ci-après dans la partie  
10 expérimentale.

La protéine hTFIIIA codée par le gène htfIIIA est donc un facteur de régulation de la transcription. En effet, la protéine hTFIIIA codée par le gène de la présente invention a un rôle biologique comme protéine se liant à l'ADN et le  
15 produit de ce gène est utile comme facteur de régulation de la transcription.

En particulier, le gène de la présente invention est exprimé dans différents tissus et joue probablement un rôle important dans l'initiation de la transcription du gène de l'ARN ribosomal 5S et dans le maintien de la stabilité de la transcrip-  
20 tion d'autres gènes impliqués notamment dans des fonctions de contrôle.

Un très grand nombre de maladies accompagnant un désordre dans le contrôle de la transcription ont récemment été mises  
25 en évidence. Ainsi on a constaté que certains produits oncogènes agissent comme des facteurs de régulation de la transcription et peuvent conduire à une cancérisation de cellules comme par exemple dans certaines leucémies ou encore que la production en trop grande quantité du facteur de  
30 régulation Hox2-4 induit une leucémie chez la souris.

Par ailleurs, dans certaines maladies héréditaires, la protéine concernée peut être en elle-même normale, la pathogénicité résultant du mécanisme de transcription du gène codant pour cette protéine. Notamment, de nombreuses maladies  
35 héréditaires montrent une anomalie de la quantité de protéines synthétisées ce qui est probablement du à un désordre au niveau de la synthèse protéique pouvant notamment mettre en jeu le gène htfIIIA et la protéine codée en tant

que facteurs impliqués dans le contrôle de la transcription de l'ARN 5S.

Le gène de la présente invention peut ainsi être utilisé pour la recherche d'anomalies au niveau de la transcription  
5 des gènes et notamment pour l'identification de maladies héréditaires ou pour l'étude de maladies mettant en cause les facteurs de régulation et notamment la protéine codée par htfIIIA.

Le gène de la présente invention peut également être  
10 utilisé pour le traitement de certaines maladies à travers le contrôle de la transcription ou dans l'analyse de la pathogénie de ces maladies.

La présente invention prévoit donc une utilisation du gène htfIIIA de la présente invention et de la protéine  
15 hTFIIIA de la présente invention, notamment pour contribuer à la compréhension du mécanisme de la transcription chez l'homme et pour contribuer également à la compréhension, au diagnostic et au traitement de maladies liées à un dérèglement du mécanisme de transcription. Ainsi htfIIIA et  
20 la protéine hTFIIIA pourraient être utilisés dans le diagnostic ou l'identification de maladies héréditaires comme certains cancers ou d'autres maladies résultant d'un contrôle anormal de la transcription. Ces facteurs peuvent également être utiles dans l'analyse des  
25 mécanismes de régulation de la transcription.

La présente invention a ainsi pour objet l'utilisation du gène du facteur de transcription humain htfIIIA ou du facteur de transcription humain codé par ce gène ainsi qu'il est défini ci-dessus pour la préparation de compositions  
30 utiles pour le diagnostic ou le traitement de maladies liées à un désordre dans le contrôle de la transcription.

De telles compositions sont préparées dans les conditions usuelles connues de l'homme du métier.

La présente invention a plus précisément pour objet  
35 l'utilisation telle que définie ci-dessus pour laquelle la maladie concernée est le cancer.

Les figures 1 à 5 ci-après présentent les illustrations suivantes.

La figure 1 représente la comparaison de la protéine hTFIIIA de la présente invention avec la protéine hTFIIIA de DREW.

La figure 2 représente la comparaison de la protéine hTFIIIA de la présente invention avec la protéine hTFIIIA de ARAKAWA.

5 La figure 3 représente la comparaison de la protéine hTFIIIA de DREW avec la protéine hTFIIIA de ARAKAWA.

La figure 4 représente la séquence htfIIIA de DREW et la protéine hTFIIIA correspondante.

10 La figure 5 représente la séquence htfIIIA de ARAKAWA et la protéine hTFIIIA correspondante.

Les séquences indiquées dans la présente invention soit : SEQ ID N°1 à SEQ ID N°10 sont décrites ci-après.

La partie expérimentale ci-après permet de décrire la présente invention sans toutefois la limiter.

## 15 Partie expérimentale

### Exemple 1 : clonage et séquençage du gène hTFIIIA

I) Extraction des ARN totaux provenant de la lignée cellulaire humaine RAJI (RNA Plus, BIOPROBE)

20 La lignée cellulaire humaine RAJI a été choisie comme source d'ARN totaux. Les cellules RAJI utilisées ont été cultivées dans les conditions usuelles de culture de cette lignée que connaît l'homme du métier.

Pour extraire les ARN totaux de ces cellules, on a procédé selon un protocole usuel en utilisant la solution

25 d'extraction commerciale RNA Plus ® (BIOPROBE SYSTEMS).

On a procédé de la manière suivante :

#### a) homogénéisation :

Les cellules cultivées en suspension sont sédimentées sans être lavées au préalable pour éviter les risques de

30 dégradation des ARNm puis sont lysées par addition de la solution d'extraction du kit RNA plus ® à raison de 6 ml pour 10<sup>7</sup> cellules. Les échantillons obtenus d'homogénat peuvent être stockés à - 70 °C.

#### b) extraction de l'ARN :

35 Après homogénéisation, on laisse l'homogénat obtenu en a) ci-dessus à 4°C pendant 5 minutes afin de permettre la complète dissociation des complexes nucléoprotéiques ; On ajoute ensuite 0.2 ml de chloroforme pour 1 ml de la solution de RNA

Plus ® ajoutée ci-dessus en a), on agite vigoureusement pendant 15 secondes et on laisse reposer dans la glace pendant 5 minutes. On centrifuge, à 12000 g et à 4°C, pendant 15 minutes.

- 5 Il se forme alors deux phases bien visibles : l'ADN et les protéines sont retrouvés dans la phase organique (phase inférieure) et à l'interface. Les ARN sont dans la phase aqueuse (phase supérieure) qui représente environ 40 à 50 % du volume total.

10 c) Précipitation de l'ARN

On transfère la phase aqueuse obtenue en b) dans un tube neuf, on ajoute un volume d'isopropanol, et on place l'échantillon, à 4°C pendant 15 minutes. On centrifuge l'échantillon pendant 15 minutes à 4°C et 12000g. On obtient  
15 un précipité qui forme un culot blanc-jaunâtre au fond du tube.

d) Lavage de l'ARN

On élimine le surnageant de la solution obtenue en c) puis on lave le culot avec une solution d'éthanol à 75% en utilisant  
20 au moins 0.8 ml d'éthanol pour 50 à 100 microgrammes d'ARN. On mélange (vortex), on centrifuge 10 minutes à 7500 g à 4°C et sèche sous vide. L'ARN obtenu est alors repris dans 60 microlitres de Tris 10 mM EDTA 1 mM pH=7,5.

II) Synthèse d'ADNc

25 a) Réactifs utilisés

Le kit commercial Gene Amp® RNA PCR Kit (Perkin Elmer) a été utilisé pour cette synthèse d'ADNc.

Par l'utilisation de ce kit, on obtient d'abord la réverse transcription de l'ARN en ADNc par la réverse transcriptase  
30 MuLV (Murine Leukemia Virus). Un inhibiteur de la RNase isolée à partir de placenta humain est inclus afin d'inhiber certaines RNases de mammifères. Les fragments d'ADNc sont amplifiés par réaction de polymérisation en chaîne (PCR). L'enzyme utilisée pour cette réaction est la pfu polymerase  
35 (Stratagène).

L'expression dNTP désigne les nucléotides dGTP, dATP, dTTP et dCTP.



L'expression PCR Buffer désigne la solution renfermant 500 mM KCl et 100 mM HCl à pH 8.3.

L'expression Oligod(T)16 désigne une séquence nucléotidique constituée de 16 nucléotides dTTP.

- 5 Des oligonucléotides sont utilisés comme amorces dans la technique décrite ci-après.

Les concentrations indiquées ci-après représentent les concentrations finales dans le milieu réactionnel.

b) Synthèse de l'ADNc par réverse transcription

- 10 2 microlitres des ARN totaux (1 microgramme) obtenus ci-dessus en 1)d) sont préincubés à 65°C pendant 5 minutes. On ajoute ensuite 8 microlitres de la solution réactionnelle suivante : 5mM MgCl<sub>2</sub>, 1xPCR buffer, 1 mM de chaque dNTP, 5% de DMSO, 1 U/microlitre d'inhibiteur de RNase, 2.5 U/microlitre de réverse transcriptase MuLV, 2.5 microlitres de oligo(dT)16. On incube alors à 42°C pendant une heure, puis à 99°C pendant 5 minutes puis à 5°C pendant 5 minutes.

III) Amplification par PCR, clonage et séquençage des séquences nucléotidiques 3' et 5'

20 a) Conditions réactionnelles

La bactérie Escherichia coli (E. coli) XL1- Blue type K12 (Stratagène) a été utilisée pour la préparation des plasmides de la présente invention.

- 25 La croissance de cette bactérie a été effectuée selon les conditions usuelles en milieu liquide LB qui renferme 10g de bactotryptone, 5g d'extrait de levure et 10 g de NaCl pour un litre d'eau et qui renferme également 100 microg/ml d'ampicilline (SIGMA).

- 30 La colonie a été prélevée sur milieu solide LB + agar + ampicilline puis cultivée dans 100 ml de milieu LB et incubée jusqu'à DO (600nm) = 0.8.

L'incubation a été effectuée à 37°C sous atmosphère normale et agitation à 225 rpm.

- 35 La viabilité de la souche est vérifiée lorsque la souche pousse sur milieu LB + ampicilline à 100 microg/ml, l'insert renfermant un gène de résistance à l'ampicilline bla.

On peut noter qu'un gène de résistance à l'ampicilline bla fait partie du vecteur du kit (TA cloning Kit - Invitrogen)

dans lequel sont clonés les fragments de htfIIIA. Ainsi, la sélection des souches renfermant les plasmides contenant le gène htfIIIA de la présente invention peut être opérée par la culture des souches dans ce milieu qui renfermant de l'ampicilline (100 microg/ml), un tel milieu permettant la survie uniquement des souches qui renferment le gène de résistance à l'ampicilline et ainsi uniquement des souches qui renferment le gène htfIIIA de la présente invention.

Pour la conservation des souches obtenues, 15 % de glycérol sont ajoutés au milieu de culture : les cultures sont donc conservées dans le milieu de suspension LB +100 microgrammes/ml d'ampicilline + 15 % de glycérol à la concentration bactérienne de DO (600nm) = 0.8 sous forme d'aliquots en cryotubes de 1ml par tube.

Pour le séquençage, l'ADN plasmidique de plusieurs bactéries issues de chacun des clonages indiqués ci-après est préparé en utilisant un kit commercial (Qiagen Plasmids kit). Les fragments correspondant à la séquence codante de htfIIIA sont séquencés sur les deux brins suivant les techniques classiques connues de l'homme du métier (utilisation du séquenceur ABI 377 XL, Perkin Elmer).

b) Amplification par PCR, et clonage des séquences nucléotidiques 3' et 5'

1) Amplification et clonage de la séquence nucléotidique 3' Deux amorces d'amplification (primers) ont été choisies d'après la séquence publiée HTFIIIA de ARAKAWA. Ces amorces OLT3 ou TFIIIA3'SmaI et OLT5.2 sont nommées respectivement SEQ ID N°10 et SEQ ID N°6.

Ces oligonucléotides sont choisis dans la séquence hTFIIIA publiée de ARAKAWA (figure 5) et sont synthétisés selon les méthodes classiques connues de l'homme du métier.

L'oligonucléotide TFIIIA3'SmaI introduit un site de restriction SmaI en aval de la séquence codante. Ce site permettra la fusion de la séquence nucléotidique 3' de htfIIIA avec une séquence codant pour le peptide TAG hémagglutinine.

Ainsi, le peptide résultant de l'expression de la séquence clonée comprendra donc à la fois la séquence de htfIIIA de la

présente invention et celle du TAG HA et pourra ainsi être détecté en analyse par Western selon les techniques usuelles connues de l'homme du métier.

On procède de la façon suivante : 2 microlitres d'ADNc obtenu  
5 ci-dessus en II)b) sont ajoutés à 50 microlitres de la solution réactionnelle suivante: 2mM MgCl<sub>2</sub>, 1xPCR buffer, 200 nanogrammes/ml de chaque dNTP, les primers TFIIIA3'SMAI et OLT5.2 à raison de 0.15 micromoles/l pour chacun, 5 % de DMSO et 2.5 U AmpliTaq ADN polymérase.

10 L'ADNc est ainsi soumis à 30 cycles PCR d'abord à 94°C , pendant une minute puis à 65°C pendant 1 minute puis à 72°C pendant 1 minute.

Les produits amplifiés par PCR ainsi obtenus sont donc des fragments 3' d'environ 970 paires de bases.

15 Les fragments 3' obtenus ci-dessus sont clonés dans le vecteur pCRII en utilisant le TA cloning Kit (Invitrogen). Le plasmide ainsi obtenu est nommé 5.2 Raji 2.9.

Ce plasmide est transféré dans la souche E. coli XL1 Blue. On obtient donc ainsi la souche E. coli transformée par le

20 plasmide 5.2 Raji 2.9.

2) Amplification et clonage de la séquence nucléotidique 5'  
La portion 5' du gène htFIIIA de la présente invention a été isolée par la technique dite de PCR ancrée 5' en utilisant un kit commercial (5'RACE System, Rapid Amplification of ADNc  
25 Ends, GIBCO BRL).

Deux amorces d'amplification (primers) ont été choisies dans la séquence publiée htFIIIA de ARAKAWA (cf. figure 5).

Ces amorces TFIIIAPCR5' et TFIIIA SEQ2 sont nommées respectivement SEQ ID N°8 et SEQ ID N°7.

30 Une chaîne homopolymérique est ajoutée à l'extrémité 3' de l'ADNc en utilisant du dATP et la Terminal desoxynucléotidyl transférase (TdT): 10 microlitres d'ADNc obtenu ci-dessus en II) b) sont incubés à 37°C pendant 10 minutes dans la solution réactionnelle 1 X tailing buffer (solution du Kit  
35 commercial)et 0.2 mM de dATP et TdT. La TdT est inactivée pendant 10 minutes à 65°C et la réaction est ensuite mise à 4°C.

La réaction est alors directement amplifiée par PCR : 10

microlitres de la réaction TdT sont ajoutés à 50 microlitres de solution réactionnelle PCR soit 1.5 mM de MgCl<sub>2</sub>, 1xPCR buffer, 200 nanomoles/ml de chaque dNTP, les primers UAP et TFIIIA PCR5' à raison de 0.2 micromoles/l pour chacun, 5% de DMSO et 2.5 U AmpliTaq ADNpolymérase.

Le primer UAP est un oligonucléotide fourni avec le kit commercial.

L'ADNc est ainsi soumis à 30 cycles PCR d'abord à 94°C pendant une minute, puis à 65°C pendant 1 minute puis à 72°C pendant 1 minute.

Les produits amplifiés par cette première PCR soit PCR1 sont soumis à une seconde réaction d'amplification par PCR en utilisant le primer UAP et un primer spécifique TFIIIA SEQ 2. On procède de la façon suivante : 5 microlitres de PCR1 sont ajoutés à 50 microlitres de la solution réactionnelle de PCR indiquée ci-dessous (1.5 mM de MgCl<sub>2</sub>, 1xPCR buffer, 200 micromoles/l de chaque dNTP, les primers UAP et TFIIIA SEQ2 à raison de 0.2 micromoles/l pour chacun, 5% de DMSO et 2.5 U AmpliTaq ADN polymérase.

L'ADN est alors soumis à 30 cycles PCR d'abord à 94°C pendant une minute, puis à 65°C pendant 1 minute puis à 72°C pendant 1 minute.

Les produits amplifiés par cette seconde PCR soit PCR2 sont purifiés sur gel d'agarose. Des fragments 5' d'environ 380 paires de bases sont ainsi isolés.

Les fragments 5' obtenus ci-dessus sont alors clonés dans le vecteur pCRII en utilisant le TA cloning Kit (Invitrogen).

Le plasmide ainsi obtenu est nommé ADNc-DMSO-3

Ce plasmide est transféré dans la souche E. coli XL1 Blue.

On obtient donc ainsi la souche E. coli transformée par le plasmide ADNc-DMSO-3.

3) Vérification de la séquence 5' par amplification d'ADN génomique - Construction du plasmide 5 geno-3

On extrait de l'ADN génomique humain à partir de cellules humaines de foie selon les méthodes usuelles connues de l'homme du métier.

On procède à une amplification par PCR de l'ADN génomique humain de la façon suivante :

2 microgrammes d'ADN génomique humain obtenu comme indiqué ci-dessus sont ajoutés à 100 microlitres de la solution réactionnelle de PCR suivante : 2mM MgCl<sub>2</sub>, 1 x native Pfu ADN polymerase buffer, 200 nanogrammes/ml de chaque dNTP, les  
 5 primers OLT5 et TFIIIASEQ2 à raison de 0.15 micromoles/l pour chacun, 5% de DMSO et 5 U pfu polymérase.  
 OLT5 et TFIIIA SEQ2 sont nommés respectivement SEQ ID N°5 et SEQ ID N°7.

Le milieu réactionnel est ainsi soumis à 30 cycles PCR  
 10 d'abord à 94°C pendant une minute, puis à 60°C pendant 1 minute puis à 72°C pendant 1 minute.

Les produits amplifiés par PCR ainsi obtenus sont des fragments d'ADN d'environ 360 paires de bases.

Les fragments ainsi obtenus sont clonés dans le vecteur PCR-  
 15 Script en utilisant le pCR-Script SK(+) Cloning kit (Stratagène).

Le plasmide ainsi obtenu est nommé 5 geno-3.

Ce plasmide est transféré dans la souche E. coli XL1 Blue. On obtient donc ainsi la souche E. coli transformée par le  
 20 plasmide 5 geno-3.

4) Clonage du gène htFIIIA selon la présente invention - Construction du plasmide pYX TFIIIALHA

On reconstitue la séquence codante htFIIIA complète par l'assemblage des deux fragments 3' et 5' obtenus ci-dessus en  
 25 III) b)1) et III)b)3)

Un site de restriction Hind III localisé sur chacun des fragments 3' et 5' obtenus ci-dessus permet de reconstituer la séquence complète.

Le plasmide 5 geno-3 obtenu ci-dessus en III)b)3) est digéré  
 30 par les enzymes de restriction EcoRI et HindIII. Le site EcoRI est situé à 11 nucléotides en amont de la séquence codante.

On obtient après purification sur gel d'agarose des fragments d'environ 350 paires de bases.

35 On procède alors à la ligation avec le vecteur pYX/EcoRI + HindIII et obtient le vecteur pYXTFIIIA5'.

On procède ensuite à l'addition du fragment 3' sur le fragment 5' : le plasmide 5.2 Raji 2.9 obtenu ci-dessus en

III) b)1), est digéré par les enzymes de restriction HindIII et SmaI.

Après purification sur gel d'agarose, on obtient un fragment d'environ 930 paires de bases. Ce fragment est inséré dans le  
5 plasmide pYXTFIIIA5' obtenu ci-dessus préalablement digéré par les enzymes de restriction SmaI et HindIII.

On obtient ainsi le plasmide pYXTFIIIALHA qui contient donc le gène hTFIIIA de la présente invention.

**Exemple 2** : Construction de la souche XL1 Blue/pYX TFIIIALHA

10 Pour la préparation de la souche XL1-Blue/ pYX TFIIIALHA, on procède selon les techniques connues de l'homme du métier (ref ci-dessus : Sambrook, Fritsh et Maniatis) à partir de la souche E. Coli XL1- Blue type K12 (Stratagène) et introduit le plasmide pYX TFIIIALHA obtenu ci-dessus à l'exemple 1.

15 **Exemple 3** : Construction du plasmide pBS-tfC2LHA

On utilise le vecteur pBS-SK+ (Stratagène) dans lequel on intègre un insert codant pour le gène htfIIIA de la présente invention. On procède comme suit : le plasmide pYXTFIIIALHA obtenu ci-dessus à l'exemple 1 est digéré par l'enzyme de  
20 restriction EcoRI, cette extrémité est remplie à l'aide de l'ADN Polymerase (fragment de Klenow) et en présence de dNTP. Ce plasmide est ensuite digéré par Nhe I et le fragment correspondant à la séquence de htfIIIA selon la présente invention est purifié. Ce fragment est inséré dans le vecteur  
25 pBS-SK+ préparé comme suit : le vecteur est digéré par EcoRI, ce site est rempli à l'aide de l'ADN polymérase puis digéré par XbaI.

On obtient ainsi le plasmide pBS-tfC2LHA.

**Exemple 4** : Construction de la souche XL1-Blue/pBS-tfC2LHA

30 Pour la préparation de la souche XL1-Blue/pBS-tfC2LHA, on procède selon les techniques connues de l'homme du métier à partir de la souche E. Coli XL1- Blue type K12 (Stratagène) et introduit le plasmide pBS-tfC2LHA obtenu ci-dessus à l'exemple 3.

35 Un échantillon de la souche obtenue soit E. Coli XL1- Blue type K12 (Stratagène) contenant le vecteur pBS-SK+ (Stratagène) avec un insert codant pour tfC2 (partie codante ADNc contenant la région codante de htfIIIA) soit XL1-

Blue/pBS-tfC2LHA a été déposé à l'Institut Pasteur 25, rue du Docteur ROUX Paris 75015 à la CNCM le 15 Septembre 1998 sous le numéro I-2071.

**Exemple 5** : Identification du codon initiateur de la synthèse

5 protéique.

La purification de la protéine hTFIIIA a été décrite par Moorefield et al (1994) [référence : the Journal of Biological Chemistry, Vol.269, N°33, pp.20857-20865, 1994, Purification and Characterization of Human Transcription Factor IIIA, B. Moorefield and R. G. Roeder].

10 La protéine hTFIIIA identifiée par Moorefield a un poids moléculaire de 42 kDa. On peut noter que le poids moléculaire théorique de la protéine hTFIIIA codée par la séquence htfIIIA de ARAKAWA est de 47 kDa.

15 La synthèse protéique est généralement initiée au niveau d'un codon ATG. Cependant la séquence codante de htfIIIA de la présente invention ne contient pas de codon ATG en phase. Il a été démontré que des codons différents de ATG, en particulier les codons CTG ou GTG sont des codons d'initia-  
20 tion de la traduction dans des transcrits cellulaires naturels.

Par des techniques connues de l'homme du métier telles que des expériences de traduction in vitro de la séquence htfIIIA selon la présente invention obtenue ci-dessus à l'exemple 1  
25 et par des tests d'expression dans des cellules de mammifères comme des cellules Cos, le codon initiateur de la synthèse protéique de hTFIIIA selon la présente invention a été mis en évidence.

Dans le cadre de la présente invention, il a ainsi été mis en  
30 évidence que le codon initiateur de la synthèse protéique de hTFIIIA selon la présente invention est le codon CTG qui se trouve à la position 176-178 de SEQ ID N°3.

**Analyse des résultats**

L'analyse des résultats obtenus par les préparations des  
35 exemples indiqués ci-dessus fait apparaître les points suivants relatifs à la séquence codante de htfIIIA :

- en 3' (ci-dessus en III)b)1)) l'essentiel de la séquence isolée dans la présente demande correspond à la séquence htfIIIA de DREW

5    - en 5' (ci-dessus en III)b)3)) la séquence la plus longue des fragments obtenus par la préparation décrite ci-dessus en III)b)3) débute en position 20 de la séquence htfIIIA de ARAKAWA et fait apparaître l'insertion d'un nucléotide en position 127 de la séquence htfIIIA de ARAKAWA.

10    Les résultats obtenus par les préparations décrites ci-dessus de htfIIIA selon la présente invention confirment qu'un nucléotide en position 127 omis dans la séquence de ARAKAWA existe bien dans le gène humain htfIIIA.



## REVENDICATIONS

- 1) Séquence d'ADN du gène htFIIIA codant pour une protéine ayant la fonction biologique du facteur de transcription humain hTFIIIA.
- 5 2) Séquence d'ADN du gène htFIIIA du facteur de transcription humain hTFIIIA selon la revendication 1 codant pour la séquence d'acides aminés SEQ ID N°2.
- 3) Séquence d'ADN du gène htFIIIA selon la revendication 1 ou 2 contenant la séquence de nucléotides SEQ ID N°3
- 10 4) Séquence d'ADN du gène htFIIIA selon les revendications 1 à 3 contenant la séquence de nucléotides SEQ ID N°4.
- 5) Séquence d'ADN selon la revendication 4 ayant la séquence commençant au nucléotide 176 et se terminant au nucléotide 1270 de SEQ ID N°3.
- 15 6) Séquence d'ADN codant pour le facteur de transcription humain hTFIIIA selon les revendications 1 à 5 ainsi que les séquences d'ADN qui hybrident avec celle-ci et/ou présentent des homologies significatives avec cette séquence ou des fragments de celle-ci et qui codent pour une protéine ayant
- 20 la même fonction.
- 7) Séquence d'ADN selon les revendications 1 à 6 comprenant des modifications introduites par suppression, insertion et/ou substitution d'au moins un nucléotide codant pour une protéine ayant la même activité biologique que le facteur de
- 25 transcription humain hTFIIIA.
- 8) Séquence d'ADN selon l'une des revendications 1 à 7 ainsi que les séquences d'ADN similaires qui ont une homologie de séquence nucléotidique d'au moins 50 % ou au moins 60 % et de préférence au moins 70 % avec ladite séquence d'ADN.
- 30 9) Séquence d'ADN selon l'une des revendications 1 à 8 ainsi que les séquences d'ADN similaires qui codent pour une protéine dont la séquence en AA a une homologie d'au moins 40 % et notamment de 45 % ou d'au moins 50 %, plutôt au moins 60 % et de préférence au moins 70 % avec la séquence
- 35 en AA codée par ladite séquence d'ADN.

- 10) Polypeptide ayant la fonction de facteur de transcription humain hTFIIIA et ayant la séquence en acides aminés SEQ ID N°2 codée par la séquence d'ADN selon l'une des revendications 1 à 9 et les analogues de ce polypeptide.
- 5 11) Procédé de préparation de la protéine recombinante hTFIIIA ayant la séquence d'acides aminés SEQ ID N°2 comprenant l'expression de la séquence d'ADN selon l'une des revendications 1 à 9 dans un hôte approprié puis l'isolement et la purification de ladite protéine
- 10 recombinante
- 12) Vecteur d'expression contenant la séquence d'ADN selon l'une des revendications 3 à 9.
- 13) Cellule hôte transformée avec un vecteur selon la revendication 12
- 15 14) Plasmide déposé à la CNCM sous le numéro I-2071
- 15) Utilisation du gène du facteur de transcription humain htfIIIA ou du facteur de transcription humain codé par ce gène selon l'une des revendications 1 à 10 pour la
- 20 préparation de compositions utiles pour le diagnostic ou le traitement de maladies liées à un désordre dans le contrôle de la transcription.
- 16) Utilisation selon la revendication 15 pour laquelle la maladie concernée est le cancer.

```

1 MDPPAVVAESVSSSLTIADAFIAAGESSAPTPPRPALPRRFICSFDPDCSAN 50
  |||||
1 ..PPAVVAESVSSSLTIADAFIAAGESSAPTPPRPALPRRFICSFDPDCSAN 48

51 YSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHTGEKPF 100
  |||||
49 YSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHTGEKPF 98

101 VCAATGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKTFKKHQQLK 150
  |||||
99 VCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKTFKKHQQLK 148

151 IHQCQHTNEPLFKCTQEGCGKHFA SP SKLRHAKAHEGYVCQKGCSFVAK 200
  |||||
149 IHQCQHTNEPLFKCTQEGCGKHFA SP SKLRHAKAHEGYVCQKGCSFVAK 198

201 TWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPERDVCRCPR 250
  |||||
199 TWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPERDVCRCPR 248

251 EGCGRYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMQSLTRHAVV 300
  |||||
249 EGCGRYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMQSLTRHAVV 298

301 HDPDKKKMKLVKKSREKRSLASHLSGYIPPKRKQGQGLSLCQNGESPNC 350
  |||||
299 HDPDKKKMKLVKKSREKRSLASHLSGYIPPKRKQGQGLSLCQNGESPNC 348

351 VEDKMLSTVAVLTLG 365
  |||||
349 VEDKMLSTVAVLTLG 363

```

FIGURE 1

```

1 .....MDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 42
:|||||||||||||||||||||||||||||||||||||||||
51 PGLGGAGALDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 100

43 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 92
|||||||||||||||||||||||||||||||||||||||||
101 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 150

93 THTGEKPFVCAATGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 142
||||||||||||| |||||||||||||||||||||||||||
151 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 200

143 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKH FASPSKLKRHAKAHEGYVCQ 192
||||||||||||| |||||||||||||||||||||||||||
201 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKH FASPSKLKRHAKAHEGYVCQ 250

193 KGCSFVAKTWTELLKXHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 242
||||||||||||| |||||||||||||||||||||||||||
251 KGCSFVAKTWTELLKXHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 300

243 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 292
||||||||||||| |||||||||||||||||||||||||||
301 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 350

293 SLTRHAVVHDPDKKKMKLKVKKKSREKRSLASHLSGYIPPKRKQGQGLSLC 342
||||||||||||| |||||||||||||||||||||||||||
351 SLTRHAVVHDPDKKKMKLKVKKKSREKREFGLSSQWIYPPPKRKQGQGLSLC 400

343 QNGESPNCVEDKMLSTVAVLTLG 365
||||||||||||| |||||||||||
401 QNGESPNCVEDKMLSTVAVLTLG 423

```

FIGURE 2

```

51 PGLGGAGALDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 100
   ||||||||||||||||||||||||||||||||||||||||||||
1  .....PPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 40

101 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 150
   ||||||||||||||||||||||||||||||||||||||||||||
41 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 90

151 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 200
   ||||||||||||||||||||||||||||||||||||||||||||
91 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 140

201 FKKHQQLKIHQCQNTNEPLFKCTQEGCGKH FASPSKLKRHAKAHEGYVCQ 250
   ||||||||||||| . ||||||||||||||||||||||||||||
141 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKH FASPSKLKRHAKAHEGYVCQ 190

251 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 300
   ||||||||||||||||||||||||||||||||||||||||||||
191 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 240

301 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 350
   ||||||||||||||||||||||||||||||||||||||||||||
241 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 290

351 SLTRHAVVHDPDKKKMKLVKKSREKREFGLSSQWIYPPKRKQGQGLSLC 400
   ||||||||||||||||||||||||| ||||||||||||
291 SLTRHAVVHDPDKKKMKLVKKSREKRS LASHLSGYIPPKRKQGQGLSLC 340

401 QNGESPNCVEDKMLSTVAVLTLG 423
   ||||||||||||||||||||
341 QNGESPNCVEDKMLSTVAVLTLG 363

```

FIGURE 3

1	CCGCCGGCCGTGGTCGCCGAGTCGGTGTGCTCCTTGACCATCGCCGACGC	50
1	P P A V V A E S V S S L T I A D A	17
51	GTTTCATTGCAGCCGGCGAGAGCTCAGCTCCGACCCCGCCGCCCCGCGC	100
18	F I A A G E S S A P T P P R P A L	34
101	TTCCCAGGAGGTTTCATCTGCTCCTTCCTGACTGCAGCGCCAATTACAGC	150
35	P R R F I C S F P D C S A N Y S	50
151	AAAGCCTGGAAGCTTGACGCGCACCTGTGCAAGCACACGGGGGAGAGACC	200
51	K A W K L D A H L C K H T G E R P	67
201	ATTTGTTTGTGACTATGAAGGGTGTGGCAAGGCCTTCATCAGGGACTACC	250
68	F V C D Y E G C G K A F I R D Y H	84
251	ATCTGAGCCGCCACATTCTGACTCACACAGGAGAAAAGCCGTTTGT	300
85	L S R H I L T H T G E K P F V C	100
301	GCAGCCAATGGCTGTGATCAAAAATTCAACACAAAATCAAACCTGAAGAA	350
101	A A N G C D Q K F N T K S N L K K	117
351	ACATTTTGAACGCAAACATGAAAATCAACAAAAACAATATATATGCAGTT	400
118	H F E R K H E N Q Q K Q Y I C S F	134
401	TTGAAGACTGTAAGAAGACCTTTAAGAAACATCAGCAGCTGAAAATCCAT	450
135	E D C K K T F K K H Q Q L K I H	150
451	CAGTGCCAGCATACCAATGAACCTCTATTCAAGTGATCCCAGGAAGGATG	500
151	Q C Q H T N E P L F K C T Q E G C	167
501	TGGGAAACACTTTGCATCACCCAGCAAGCTGAAACGACATGCCAAGGCCC	550
168	G K H F A S P S K L K R H A K A H	184
551	ACGAGGGCTATGTATGTCAAAAAGGATGTTCCCTTTGTGGCAAAAACATGG	600
185	E G Y V C Q K G C S F V A K T W	200
601	ACGGAACCTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATG	650
201	T E L L K H V R E T H K E E I L C	217

FIGURE 4

651	TGAAGTATGCCGAAAAACATTTAAACGCAAAGATTACCTTAAGCAACACA	700
218	E V C R K T F K R K D Y L K Q H M	234
701	TGAAAACTCATGCCCCAGAAAGGGATGTATGTCGCTGTCCAAGAGAAGGC	750
235	K T H A P E R D V C R C P R E G	250
751	TGTGGAAGAACCTATACAACTGTGTTTAATCTCCAAAGCCATATCCTCTC	800
251	C G R T Y T T V F N L Q S H I L S	267
801	CTTCCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCA	850
268	F H E E S R P F V C E H A G C G K	284
851	AAACATTTGCAATGAAACAAAGTCTCACTAGGCATGCTGTTGTACATGAT	900
285	T F A M K Q S L T R H A V V H D	300
901	CCTGACAAGAAGAAAATGAAGCTCAAAGTCAAAAATCTCGTGAAAAACG	950
301	P D K K K M K L K V K K S R E K R	317
951	GAGTTTGGCCTCTCATCTCAGTGGATATATCCCTCCCAAAGGAAACAAG	1000
318	S L A S H L S G Y I P P K R K Q G	334
1001	GGCAAGGCTTATCTTTGTGTCAAAACGGAGAGTCACCCAACTGTGTGGAA	1050
335	Q G L S L C Q N G E S P N C V E	350
1051	GACAAGATGCTCTCGACAGTTGCAGTACTTACCCTTGGCTAAGAACTGCA	1100
351	D K M L S T V A V L T L G *	364
1101	CTGCTTTGTTTAAAGGACTGCAGACCAAGGAGCGAGCTTTCTCTCAGAGC	1150
1151	ATGCTTTTCTTTATTAAAATTAC	1173

FIGURE 4

1	ATGCGCGATCTCCCGGAGCATGCGCAGCAGCGGCGCCGACGCGGGGCGGT	50
1	M R S S G A D A G R C	11
51	GCCTGGTGACCGCGCGCGCTCCCGGAAGTGTGCCGGCGTCGCGCGAAGGT	100
12	L V T A R A P G S V P A S R E G	27
101	TCAGCAGGGAGCCGTGGGCGGGCGCGCGGTTCCCGGCACGTGTCTCGGC	150
28	S A G S R G P G A R F P A R V S A	44
151	ACGTGGCAGCGCGCCTGGCCCTGGGCTTGGAGGCGCCGGCGCCCTGGATC	200
45	R G S A P G P G L G G A G A L D P	61
201	CGCCGGCCGTGGTCGCCGAGTCGGTGTGTCCTTGACCATCGCCGACGCG	250
62	P A V V A E S V S S L T I A D A	77
251	TTCATTGCAGCCGGCGAGAGCTCAGCTCCGACCCCGCCGCGCCCGCGCT	300
78	F I A A G E S S A P T P P R P A L	94
301	TCCCAGGAGGTTTCATCTGCTCCTTCCCTGACTGCAGCGCCAATTACAGCA	350
95	P R R F I C S F P D C S A N Y S K	111
351	AAGCCTGGAAGCTTGACGCGCACCTGTGCAAGCACACGGGGGAGAGACCA	400
112	A W K L D A H L C K H T G E R P	127
401	TTTGTGTTGTGACTATGAAGGGTGTGGCAAGGCCTTCATCAGGGACTACCA	450
128	F V C D Y E G C G K A F I R D Y H	144
451	TCTGAGCCGCCACATTCTGACTCACACAGGAGAAAAGCCGTTTGTGTTGTG	500
145	L S R H I L T H T G E K P F V C A	161
501	CAGCCAATGGCTGTGATCAAAAATTCAACACAAAATCAAACCTGAAGAAA	550
162	A N G C D Q K F N T K S N L K K	177
551	CATTTTGAACGCAAACATGAAAATCAACAAAAACAATATATATGCAGTTT	600
178	H F E R K H E N Q Q K Q Y I C S F	194
601	TGAAGACTGTAAGAAGACCTTTAAGAAACATCAGCAGCTGAAAATCCATC	650
195	E D C K K T F K K H Q Q L K I H Q	211
651	AGTGCCAGAATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGGATGT	700
212	C Q N T N E P L F K C T Q E G C	227

FIGURE 5



701	GGGAAACACTTTGCATCACCCAGCAAGCTGAAACGACATGCCAAGGCCCA	750
228	G K H F A S P S K L K R H A K A H	244
751	CGAGGGCTATGTATGTCAAAAAGGATGTTCCCTTTGTGGCAAAAACATGGA	800
245	E G Y V C Q K G C S F V A K T W T	261
801	CGGAACTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATGT	850
262	E L L K H V R E T H K E E I L C	277
851	GAAGTATGCCGAAAAACATTTAAACGCAAAGATTACCTTAAGCAACACAT	900
278	E V C R K T F K R K D Y L K Q H M	294
901	GAAAACTCATGCCCCAGAAAGGGATGTATGTCGCTGTCCAAGAGAAGGCT	950
295	K T H A P E R D V C R C P R E G C	311
951	GTGGAAGAACCTATACAACCTGTGTTTAATCTCCAAAGCCATATCCTCTCC	1000
312	G R T Y T T V F N L Q S H I L S	327
1001	TTCCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCAA	1050
328	F H E E S R P F V C E H A G C G K	344
1051	AACATTTGCAATGAAACAAAGTCTCACTAGGCATGCTGTTGTACATGATC	1100
345	T F A M K Q S L T R H A V V H D P	361
1101	CTGACAAGAAGAAAATGAAGCTCAAAGTCAAAAAATCTCGTGAAAAACGG	1150
362	D K K K M K L K V K K S R E K R	377
1151	GAGTTTGGCCTCTCATCTCAGTGGATATATCCTCCCAAAGGAAACAAGG	1200
378	E F G L S S Q W I Y P P K R K Q G	394
1201	GCAAGGCTTATCTTTGTGTCAAAACGGAGAGTCACCCAACCTGTGTGGAAG	1250
395	Q G L S L C Q N G E S P N C V E D	411
1251	ACAAGATGCTCTCGACAGTTGCAGTACTTACCCTTGGCTAAGAACTGCAC	1300
412	K M L S T V A V L T L G *	424
1301	TGCTTTGTTTAAAGGACTGCAGACCAAGGAGTCGAGCTTTCTCTCAGAGC	1350
ATGCTTTTCTTTATTAAATTACTGATGCAGAAAAAAAAAAAAAAAAAAAA		1399

FIGURE 5

## LISTAGE DE SEQUENCE

&lt;110&gt; Hoechst Marion Roussel

&lt;120&gt; Gène humain htFIIIA et la protéine codée htFIIIA.

&lt;130&gt; 9823seq

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 10

&lt;170&gt; PatentIn Vers. 2.0

&lt;210&gt; 1

&lt;211&gt; 1273

&lt;212&gt; ADN

&lt;213&gt; Human

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (176) .. (1270)

&lt;400&gt; 1

```

atgcgcagca gcggcgccga cgcgggcgcg tgccctggtga ccgcgcgcgc tcccgggaagt 60
gtgccggcggt cgcgcggaagg ttcagcaggg agccgtgggc cgggcgcgcc gggtcccggc 120
acgtgtctctg gcacgtggca gcgcgcctgg ccctggggctt ggaggcgccg gcgcc ctg 178
                                                    Met
                                                    1

gat ccg ccg gcc gtg gtc gcc gag tcg gtg tcg tcc ttg acc atc gcc 226
Asp Pro Pro Ala Val Val Ala Glu Ser Val Ser Ser Leu Thr Ile Ala
           5                10                15

gac gcg ttc att gca gcc ggc gag agc tca gct ccg acc ccg ccg cgc 274
Asp Ala Phe Ile Ala Ala Gly Glu Ser Ser Ala Pro Thr Pro Pro Arg
           20                25                30

ccc gcg ctt ccc agg agg ttc atc tgc tcc ttc cct gac tgc agc gcc 322
Pro Ala Leu Pro Arg Arg Phe Ile Cys Ser Phe Pro Asp Cys Ser Ala
           35                40                45

aat tac agc aaa gcc tgg aag ctt gac gcg cac ctg tgc aag cac acg 370
Asn Tyr Ser Lys Ala Trp Lys Leu Asp Ala His Leu Cys Lys His Thr
           50                55                60                65

ggg gag aga cca ttt gtt tgt gac tat gaa ggg tgt ggc aag gcc ttc 418
Gly Glu Arg Pro Phe Val Cys Asp Tyr Glu Gly Cys Gly Lys Ala Phe
           70                75                80

atc agg gac tac cat ctg agc cgc cac att ctg act cac aca gga gaa 466
Ile Arg Asp Tyr His Leu Ser Arg His Ile Leu Thr His Thr Gly Glu

```

85	90	95	
aag ccg ttt gtt tgt gca gcc act ggc tgt gat caa aaa ttc aac aca			514
Lys Pro Phe Val Cys Ala Ala Thr Gly Cys Asp Gln Lys Phe Asn Thr			
100	105	110	
aaa tca aac ttg aag aaa cat ttt gaa cgc aaa cat gaa aat caa caa			562
Lys Ser Asn Leu Lys Lys His Phe Glu Arg Lys His Glu Asn Gln Gln			
115	120	125	
aaa caa tat ata tgc agt ttt gaa gac tgt aag aag acc ttt aag aaa			610
Lys Gln Tyr Ile Cys Ser Phe Glu Asp Cys Lys Lys Thr Phe Lys Lys			
130	135	140	145
cat cag cag ctg aaa atc cat cag tgc cag cat acc aat gaa cct cta			658
His Gln Gln Leu Lys Ile His Gln Cys Gln His Thr Asn Glu Pro Leu			
150	155	160	
ttc aag tgt acc cag gaa gga tgt ggg aaa cac ttt gca tca ccc agc			706
Phe Lys Cys Thr Gln Glu Gly Cys Gly Lys His Phe Ala Ser Pro Ser			
165	170	175	
aag ctg aaa cga cat gcc aag gcc cac gag ggc tat gta tgt caa aaa			754
Lys Leu Lys Arg His Ala Lys Ala His Glu Gly Tyr Val Cys Gln Lys			
180	185	190	
gga tgt tcc ttt gtg gca aaa aca tgg acg gaa ctt ctg aaa cat gtg			802
Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His Val			
195	200	205	
aga gaa acc cat aaa gag gaa ata cta tgt gaa gta tgc cgg aaa aca			850
Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys Thr			
210	215	220	225
ttt aaa cgc aaa gat tac ctt aag caa cac atg aaa act cat gcc cca			898
Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala Pro			
230	235	240	
gaa agg gat gta tgt cgc tgt cca aga gaa ggc tgt gga aga acc tat			946
Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr Tyr			
245	250	255	
act act gtg ttt aat ctc caa agc cat atc ctc tcc ttc cat gag gaa			994
Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu Glu			
260	265	270	
agc cgc cct ttt gtg tgt gaa cat gct ggc tgt ggc aaa aca ttt gca			1042
Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe Ala			
275	280	285	
atg aaa caa agt ctc act agg cat gct gtt gta cat gat cct gac aag			1090
Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp Lys			
290	295	300	305
aag aaa atg aag ctc aaa gtc aaa aaa tct cgt gaa aaa cgg agt ttg			1138
Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser Leu			
310	315	320	

Lys Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His

195	200	205
Val Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys 210	215	220
Thr Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala 225	230	235 240
Pro Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr 245	250	255
Tyr Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu 260	265	270
Glu Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe 275	280	285
Ala Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp 290	295	300
Lys Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser 305	310	315 320
Leu Ala Ser His Leu Ser Gly Tyr Ile Pro Pro Lys Arg Lys Gln Gly 325	330	335
Gln Gly Leu Ser Leu Cys Gln Asn Gly Glu Ser Pro Asn Cys Val Glu 340	345	350
Asp Lys Met Leu Ser Thr Val Ala Val Leu Thr Leu Gly 355	360	365

<210> 3  
 <211> 1273  
 <212> ADN  
 <213> Human

<400> 3  
 atgcgcagca gcgccgcga cgccggcggtg cgcgcgcgc tcccgggaagt 60  
 gtgccggcgt cgcgcgaagg ttcagcaggg agccgtgggc cgccgcgcgc gggtcccggc 120  
 acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccttgga 180  
 tccgcgggcc gtggtcgccg agtcgggtgtc gtccttgacc atcgccgacg cgttcattgc 240  
 agccggcgag agtcagctc cgaccccgcc gcgccccgcg cttcccagga gggtcatctg 300  
 ctcttccct gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 360  
 caagcacacg ggggagagac catttggttg tgactatgaa ggggtgtggca aggccttcat 420  
 cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttggttg 480  
 tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aactgaaga aacattttga 540

acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 600  
 ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 660  
 caagtgtacc caggaaggat gtgggaaaca ctttgcata cccagcaagc tgaaacgaca 720  
 tgccaaggcc cagcagggct atgtatgtca aaaaggatgt tcctttgttg caaaaacatg 780  
 gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaatactat gtgaagtatg 840  
 ccggaanaa tttaaacgca aagattacct taagcaacac atgaaaactc atgccccaga 900  
 aagggatgta tgtcgctgtc caagagaagg ctgtggaaga acctatacta ctgtgtttta 960  
 tctccaaagc catatcctct ccttccatga ggaaagccgc ccttttgtgt gtgaacatgc 1020  
 tggctgtggc aaaacatttg caatgaaaca aagtctcact aggcattgtg ttgtacatga 1080  
 tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1140  
 ctctcatctc agtggatata tccctccaa aaggaaacaa gggcaaggct tatctttgtg 1200  
 tcaaacgga gagtcacca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1260  
 tacccttggc taa 1273

<210> 4  
 <211> 1213  
 <212> ADN  
 <213> Human

<400> 4  
 gtgccggcgc cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccggc 60  
 acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccttgga 120  
 tccgccggcc gtggctgcgc agtcgggtgtc gtccttgacc atcgccgacg cgttcattgc 180  
 agccggcgag agctcagctc cgaccccgcc gcgccccgcg ctteccagga ggttcattcg 240  
 ctcttccct gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 300  
 caagcacacg ggggagagac catttgtttg tgactatgaa ggggtgtggc aggccttcat 360  
 cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttgtttg 420  
 tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aacttgaaga aacattttga 480  
 acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 540  
 ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 600  
 caagtgtacc caggaaggat gtgggaaaca ctttgcata cccagcaagc tgaaacgaca 660  
 tgccaaggcc cagcagggct atgtatgtca aaaaggatgt tcctttgttg caaaaacatg 720

gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaatactat gtgaagtatg 780  
 ccggaaaaca tttaaacgca aagattacct taagcaacac atgaaaactc atgccccaga 840  
 aagggatgta tgtcgctgtc caagagaagg ctgtggaaga acctatacta ctgtgtttaa 900  
 tctccaaagc catatcctct ccttccatga ggaaagccgc ccttttgtgt gtgaacatgc 960  
 tggctgtggc aaaacatttg caatgaaaca aagtctcact aggcattgctg ttgtacatga 1020  
 tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1080  
 ctctcatctc agtggatata tccctcccaa aaggaaacaa gggcaaggct tatctttgtg 1140  
 tcaaaacgga gagtcacca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1200  
 tacccttggc taa 1213

<210> 5  
 <211> 34  
 <212> ADN  
 <213> Human

<400> 5  
 cggggtacca aaaatgcgca gcagcggcgc cgac 34

<210> 6  
 <211> 21  
 <212> ADN  
 <213> Human

<400> 6  
 tccttccttg actgcagcgc c 21

<210> 7  
 <211> 20  
 <212> ADN  
 <213> Human

<400> 7  
 tgcacagggtg cgcgtcaagc 20

<210> 8  
 <211> 20  
 <212> ADN  
 <213> Human

<400> 8  
 cacaaacaaa tggctctctcc 20

<210> 9  
 <211> 30

<212> ADN  
<213> Human

<400> 9  
cggcttagat tagccaaggg taagtactgc

30

<210> 10  
<211> 30  
<212> ADN  
<213> Human

<400> 10  
cctcccgggg ccaagggtaa gtactgcaac

30



que facteurs impliqués dans le contrôle de la transcription de l'ARN 5S.

Le gène de la présente invention peut ainsi être utilisé pour la recherche d'anomalies au niveau de la transcription  
5 des gènes et notamment pour l'identification de maladies héréditaires ou pour l'étude de maladies mettant en cause les facteurs de régulation et notamment la protéine codée par htfIIIA.

Le gène de la présente invention peut également être  
10 utilisé pour le traitement de certaines maladies à travers le contrôle de la transcription ou dans l'analyse de la pathogénie de ces maladies.

La présente invention prévoit donc une utilisation du gène htfIIIA de la présente invention et de la protéine  
15 hTFIIIA de la présente invention, notamment pour contribuer à la compréhension du mécanisme de la transcription chez l'homme et pour contribuer également à la compréhension, au diagnostic et au traitement de maladies liées à un dérèglement du mécanisme de transcription. Ainsi htfIIIA et  
20 la protéine hTFIIIA pourraient être utilisés dans le diagnostic ou l'identification de maladies héréditaires comme certains cancers ou d'autres maladies résultant d'un contrôle anormal de la transcription. Ces facteurs peuvent également être utiles dans l'analyse des mécanismes de régulation de la  
25 transcription.

La présente invention a ainsi pour objet l'utilisation de la séquence d'ADN du gène du facteur de transcription humain htfIIIA ou du polypeptide ayant la fonction de facteur de transcription humain codé par ce gène ainsi qu'il est  
30 défini ci-dessus pour la préparation de compositions utiles pour le diagnostic ou le traitement de maladies liées à un désordre dans le contrôle de la transcription. De telles compositions sont préparées dans les conditions usuelles connues de l'homme du métier.

35 La présente invention a plus précisément pour objet l'utilisation telle que définie ci-dessus pour laquelle la maladie concernée est le cancer. Les figures 1 à 5 ci-après présentent les illustrations suivantes. La figure 1 représente la comparaison de la protéine hTFIIIA

## REVENDICATIONS

- 1) Séquence d'ADN du gène htFIIIA codant pour une protéine ayant la fonction biologique du facteur de transcription humain hTFIIIA.
- 5 2) Séquence d'ADN du gène htFIIIA du facteur de transcription humain hTFIIIA selon la revendication 1 codant pour la séquence d'acides aminés SEQ ID N°2.
- 3) Séquence d'ADN du gène htFIIIA selon la revendication 1 ou 2 contenant la séquence de nucléotides SEQ ID N°3
- 10 4) Séquence d'ADN du gène htFIIIA selon les revendications 1 à 3 contenant la séquence de nucléotides SEQ ID N°4.
- 5) Séquence d'ADN selon la revendication 4 ayant la séquence commençant au nucléotide 176 et se terminant au nucléotide 1270 de SEQ ID N°3.
- 15 6) Séquence d'ADN codant pour le facteur de transcription humain hTFIIIA selon les revendications 1 à 5 ainsi que les séquences d'ADN qui hybrident avec celle-ci et/ou présentent des homologies significatives avec cette séquence ou des fragments de celle-ci et qui codent pour une protéine ayant
- 20 la même fonction.
- 7) Séquence d'ADN selon les revendications 1 à 6 comprenant des modifications introduites par suppression, insertion et/ou substitution d'au moins un nucléotide codant pour une protéine ayant la même activité biologique que le facteur de
- 25 transcription humain hTFIIIA.
- 8) Séquence d'ADN selon l'une des revendications 1 à 7 ainsi que les séquences d'ADN similaires qui ont une homologie de séquence nucléotidique d'au moins 50 % ou au moins 60 % et de préférence au moins 70 % avec ladite séquence d'ADN.
- 30 9) Séquence d'ADN selon l'une des revendications 1 à 8 ainsi que les séquences d'ADN similaires qui codent pour une protéine dont la séquence en AA a une homologie d'au moins 40 % et notamment de 45 % ou d'au moins 50 %, plutôt au moins 60 % et de préférence au moins 70 % avec la séquence
- 35 en AA codée par ladite séquence d'ADN.

- 10) Polypeptide ayant la fonction de facteur de transcription humain hTFIIIA et ayant la séquence en acides aminés SEQ ID N°2 codée par la séquence d'ADN selon l'une des revendications 1 à 9 et les analogues de ce polypeptide.
- 5 11) Procédé de préparation de la protéine recombinante hTFIIIA ayant la séquence d'acides aminés SEQ ID N°2 comprenant l'expression de la séquence d'ADN selon l'une des revendications 1 à 9 dans un hôte approprié puis l'isolement et la purification de ladite protéine
- 10 recombinante
- 12) Vecteur d'expression contenant la séquence d'ADN selon l'une des revendications 3 à 9.
- 13) Cellule hôte transformée avec un vecteur selon la revendication 12
- 15 14) Plasmide déposé à la CNCM sous le numéro I-2071
- 15) Utilisation de la séquence d'ADN du gène htfIIIA selon l'une des revendications 1 à 9 ou du polypeptide ayant la fonction de facteur de transcription humain codé par ladite séquence d'ADN et tel que défini à la revendication 10 pour
- 20 la préparation de compositions utiles pour le diagnostic ou le traitement de maladies liées à un désordre dans le contrôle de la transcription.
- 16) Utilisation selon la revendication 15 pour laquelle la maladie concernée est le cancer.

```

1 MDPPAVVAESVSSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSAN 50
  |||
1 ..PPAVVAESVSSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSAN 48
  |||
51 YSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHTGKPF 100
  |||
49 YSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHTGKPF 98
  |||
101 VCAATGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKTFKKHQQLK 150
  |||
99 VCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKTFKKHQQLK 148
  |||
151 IHQCQHTNEPLFKCTQEGCGKHFA SP SKLRHAKAHEGYVCQKGCSFVAK 200
  |||
149 IHQCQHTNEPLFKCTQEGCGKHFA SP SKLRHAKAHEGYVCQKGCSFVAK 198
  |||
201 TWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPERDVCRCPR 250
  |||
199 TWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPERDVCRCPR 248
  |||
251 EGCGRITYTTVFNLQSHILSFHEESRPVFCEHAGCGKTFAMKQSLTRHAVV 300
  |||
249 EGCGRITYTTVFNLQSHILSFHEESRPVFCEHAGCGKTFAMKQSLTRHAVV 298
  |||
301 HDPDKKKMKLVKKSREKRSLASHLSGYIPPKRKQGQGLSLCQNGESPNC 350
  |||
299 HDPDKKKMKLVKKSREKRSLASHLSGYIPPKRKQGQGLSLCQNGESPNC 348
  |||
351 VEDKMLSTVAVLTLG 365
  |||
349 VEDKMLSTVAVLTLG 363

```

FIGURE 1

```

1 .....MDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 42
:|||||||||||||||||||||||||||||||||||||||||
51 PGLGGAGALDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 100

43 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 92
|||||||||||||||||||||||||||||||||||||||||
101 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 150

93 THTGEKPFVCAATGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 142
|||||||||||||||||||||||||||||||||||||||||
151 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 200

143 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLRHAKAHEGYVCQ 192
|||||||||||||||||||||||||||||||||||||||||
201 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLRHAKAHEGYVCQ 250

193 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 242
|||||||||||||||||||||||||||||||||||||||||
251 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 300

243 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 292
|||||||||||||||||||||||||||||||||||||||||
301 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 350

293 SLTRHAVVHDPDKKKMKLVKKSREKRSLASHLSGYIPPKRKQGQGLSLC 342
||||||||||||||||||||||||| |||||||||||
351 SLTRHAVVHDPDKKKMKLVKKSREKREFGLSSQWIYPPPKRKQGQGLSLC 400

343 QNGESPNCVEDKMLSTVAVLTLG 365
|||||||||||||
401 QNGESPNCVEDKMLSTVAVLTLG 423

```

FIGURE 2

```

51 PGLGGAGALDPPAVVAESVSSLTIAADAFIAAGESSAPTPPRPALPRRFIC 100
   ||||||||||||||||||||||||||||||||||||||||||||
1  .....PPAVVAESVSSLTIAADAFIAAGESSAPTPPRPALPRRFIC 40
101 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 150
   ||||||||||||||||||||||||||||||||||||||||||||
41 SFPDCSANYSKAWKLD AHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 90
151 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQOKQYICSFEDCKKT 200
   ||||||||||||||||||||||||||||||||||||||||||||
91 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQOKQYICSFEDCKKT 140
201 FKKHQQLKIHQCQNTNEPLFKCTQEGCGKH FASPSKLKRHAKAHEGYVCQ 250
   ||||||||||||||||||||||||||||||||||||||||||||
141 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKH FASPSKLKRHAKAHEGYVCQ 190
251 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTTHAPE 300
   ||||||||||||||||||||||||||||||||||||||||||||
191 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTTHAPE 240
301 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 350
   ||||||||||||||||||||||||||||||||||||||||||||
241 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 290
351 SLTRHAVVHDPDKKKMKLVKKSREKREFGLSSQWIYPPKRKQGQGLSLC 400
   ||||||||||||||||||||||||||||| |||||||||||||
291 SLTRHAVVHDPDKKKMKLVKKSREKRSLASHLSGYIPPKRKQGQGLSLC 340
401 QNGESPNCVEDKMLSTVAVLTLG 423
   |||||||||||||||||||||||||
341 QNGESPNCVEDKMLSTVAVLTLG 363

```

FIGURE 3

1	CCGCCGGCCGTGGTCGCCGAGTCGGTGTCTCCTTGACCATCGCCGACGC	50
1	P P A V V A E S V S S L T I A D A	17
51	GTTTCATTGCAGCCGGCGAGAGCTCAGCTCCGACCCCGCCGCCCCGCGC	100
18	F I A A G E S S A P T P P R P A L	34
101	TTCCCAGGAGGTTTCATCTGCTCCTTCCCTGACTGCAGCGCCAATTACAGC	150
35	P R R F I C S F P D C S A N Y S	50
151	AAAGCCTGGAAGCTTGACGCGCACCTGTGCAAGCACACGGGGGAGAGAC	200
51	K A W K L D A H L C K H T G E R P	67
201	ATTTGTTTGTGACTATGAAGGGTGTGGCAAGGCCTTCATCAGGGACTACC	250
68	F V C D Y E G C G K A F I R D Y H	84
251	ATCTGAGCCGCCACATTCTGACTCACACAGGAGAAAAGCCGTTTGTTTGT	300
85	L S R H I L T H T G E K P F V C	100
301	GCAGCCAATGGCTGTGATCAAAAATTCAACACAAAATCAAACCTTGAAGAA	350
101	A A N G C D Q K F N T K S N L K K	117
351	ACATTTTGAACGCAAACATGAAAATCAACAAAAACAATATATATGCAGTT	400
118	H F E R K H E N Q Q K Q Y I C S F	134
401	TTGAAGACTGTAAGAAGACCTTTAAGAAACATCAGCAGCTGAAAATCCAT	450
135	E D C K K T F K K H Q Q L K I H	150
451	CAGTGCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGGATG	500
151	Q C Q H T N E P L F K C T Q E G C	167
501	TGGGAAACACTTTGCATCACCCAGCAAGCTGAAACGACATGCCAAGGCCC	550
168	G K H F A S P S K L K R H A K A H	184
551	ACGAGGGCTATGTATGTCAAAAAGGATGTTTCCTTTGTGGCAAAAACATGG	600
185	E G Y V C Q K G C S F V A K T W	200
601	ACGGAACCTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATG	650
201	T E L L K H V R E T H K E E I L C	217

FIGURE 4

651	TGAAGTATGCCGGAACATTTAAACGCAAAGATTACCTTAAGCAACACA	700
218	E V C R K T F K R K D Y L K Q H M	234
701	TGAAAACTCATGCCCCAGAAAGGGATGTATGTCGCTGTCCAAGAGAAGGC	750
235	K T H A P E R D V C R C P R E G	250
751	TGTGGAAGAACCTATACAACCTGTGTTTAATCTCCAAAGCCATATCCTCTC	800
251	C G R T Y T T V F N L Q S H I L S	267
801	CTTCCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCA	850
268	F H E E S R P F V C E H A G C G K	284
851	AAACATTTGCAATGAAACAAAGTCTCACTAGGCATGCTGTTGTACATGAT	900
285	T F A M K Q S L T R H A V V H D	300
901	CCTGACAAGAAGAAAATGAAGCTCAAAGTCAAAAAATCTCGTGAAAAACG	950
301	P D K K K M K L K V K K S R E K R	317
951	GAGTTTGGCCTCTCATCTCAGTGGATATATCCCTCCCAAAGGAAACAAG	1000
318	S L A S H L S G Y I P P K R K Q G	334
1001	GGCAAGGCTTATCTTTGTGTCAAACGGAGAGTCACCCAACCTGTGTGGAA	1050
335	Q G L S L C Q N G E S P N C V E	350
1051	GACAAGATGCTCTCGACAGTTGCAGTACTTACCCTTGGCTAAGAACTGCA	1100
351	D K M L S T V A V L T L G *	364
1101	CTGCTTTGTTTAAAGGACTGCAGACCAAGGAGCGAGCTTTCTCTCAGAGC	1150
1151	ATGCTTTTCTTTATTAAATTAC	1173

FIGURE 4



1	ATGCGCGATCTCCCGGAGCATGCGCAGCAGCGGCGCCGACGCGGGGCGGT	50
1	M R S S G A D A G R C	11
51	GCCTGGTGACCGCGCGCTCCCGGAAGTGTGCCGGCGTCGCGCGAAGGT	100
12	L V T A R A P G S V P A S R E G	27
101	TCAGCAGGGAGCCGTGGGCGGGCGCGCGGTTCCTGGCACGTGTCTCGGC	150
28	S A G S R G P G A R F P A R V S A	44
151	ACGTGGCAGCGCGCCTGGCCCTGGGCTTGGAGGCGCCGGCGCCCTGGATC	200
45	R G S A P G P G L G G A G A L D P	61
201	CGCCGGCCGTGGTCGCCGAGTCGGTGTCTGCTCCTTGACCATCGCCGACGCG	250
62	P A V V A E S V S S L T I A D A	77
251	TTCATTGCAGCCGGCGAGAGCTCAGCTCCGACCCGCGCGCCCCGCGCT	300
78	F I A A G E S S A P T P P R P A L	94
301	TCCCAGGAGGTTTCATCTGCTCCTTCCCTGACTGCAGCGCCAATTACAGCA	350
95	P R R F I C S F P D C S A N Y S K	111
351	AAGCCTGGAAGCTTGACGCGCACCTGTGCAAGCACACGGGGGAGAGACCA	400
112	A W K L D A H L C K H T G E R P	127
401	TTTGTGTTGTGACTATGAAGGGTGTGGCAAGGCCTTCATCAGGGACTACCA	450
128	F V C D Y E G C G K A F I R D Y H	144
451	TCTGAGCCGCCACATTCTGACTCACACAGGAGAAAAGCCGTTTGTGTTGTG	500
145	L S R H I L T H T G E K P F V C A	161
501	CAGCCAATGGCTGTGATCAAAAATTCAACACAAAATCAAATCTGAAGAAA	550
162	A N G C D Q K F N T K S N L K K	177
551	CATTTTGAACGCAACATGAAAATCAACAAAAACAATATATATGCAGTTT	600
178	H F E R K H E N Q Q K Q Y I C S F	194
601	TGAAGACTGTAAGAAGACCTTTAAGAAACATCAGCAGCTGAAAATCCATC	650
195	E D C K K T F K K H Q Q L K I H Q	211
651	AGTGCCAGAATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGGATGT	700
212	C Q N T N E P L F K C T Q E G C	227

FIGURE 5

701	GGGAAACACTTTGCATCACCCAGCAAGCTGAAACGACATGCCAAGGCCCA	750
228	G K H F A S P S K L K R H A K A H	244
751	CGAGGGCTATGTATGTCAAAAAGGATGTTCCCTTTGTGGCAAAAACATGGA	800
245	E G Y V C Q K G C S F V A K T W T	261
801	CGGAACTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATGT	850
262	E L L K H V R E T H K E E I L C	277
851	GAAGTATGCCGGAAAACATTTAAACGCAAAGATTACCTTAAGCAACACAT	900
278	E V C R K T F K R K D Y L K Q H M	294
901	GAAAACTCATGCCCCAGAAAGGGATGTATGTCGCTGTCCAAGAGAAGGCT	950
295	K T H A P E R D V C R C P R E G C	311
951	GTGGAAGAACCTATACAACCTGTGTTTAATCTCCAAAGCCATATCCTCTCC	1000
312	G R T Y T T V F N L Q S H I L S	327
1001	TTCCATGAGGAAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCAA	1050
328	F H E E S R P F V C E H A G C G K	344
1051	AACATTTGCAATGAAACAAAGTCTCACTAGGCATGCTGTTGTACATGATC	1100
345	T F A M K Q S L T R H A V V H D P	361
1101	CTGACAAGAAGAAAATGAAGCTCAAAGTCAAAAAATCTCGTGAAAAACGG	1150
362	D K K K M K L K V K K S R E K R	377
1151	GAGTTTGGCCTCTCATCTCAGTGGATATATCCTCCCAAAGGAAACAAGG	1200
378	E F G L S S Q W I Y P P K R K Q G	394
1201	GCAAGGCTTATCTTTGTGTCAAAACGGAGAGTCACCCAAGTGTGTGGAAG	1250
395	Q G L S L C Q N G E S P N C V E D	411
1251	ACAAGATGCTCTCGACAGTTGCAGTACTTACCCTTGGCTAAGAACTGCAC	1300
412	K M L S T V A V L T L G *	424
1301	TGCTTTGTTTAAAGGACTGCAGACCAAGGAGTCGAGCTTTCTCTCAGAGC	1350

ATGCTTTTCTTTATTAAATTACTGATGCAGAAAAAAAAAAAAAAAAAAAA 1399

FIGURE 5

## LISTAGE DE SEQUENCE

&lt;110&gt; Hoechst Marion Roussel

&lt;120&gt; Gène humain htfIIIA et la protéine codée htfIIIA.

&lt;130&gt; 9823seq

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 10

&lt;170&gt; PatentIn Vers. 2.0

&lt;210&gt; 1

&lt;211&gt; 1273

&lt;212&gt; ADN

&lt;213&gt; Human

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (176)..(1270)

&lt;400&gt; 1

```

atgcgcgagca gcggcgccga cgcggggcgg tgccctggtga ccgcgcgcgc tcccgggaagt 60
gtgccggcgt cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc gggtcccgcc 120
acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgcc ctg 178
                                         Met
                                         1

gat ccg ccg gcc gtg gtc gcc gag tcg gtg tcg tcc ttg acc atc gcc 226
Asp Pro Pro Ala Val Val Ala Glu Ser Val Ser Ser Leu Thr Ile Ala
          5                      10                      15

gac gcg ttc att gca gcc ggc gag agc tca gct ccg acc ccg ccg cgc 274
Asp Ala Phe Ile Ala Ala Gly Glu Ser Ser Ala Pro Thr Pro Pro Arg
          20                      25                      30

ccc gcg ctt ccc agg agg ttc atc tgc tcc ttc cct gac tgc agc gcc 322
Pro Ala Leu Pro Arg Arg Phe Ile Cys Ser Phe Pro Asp Cys Ser Ala
          35                      40                      45

aat tac agc aaa gcc tgg aag ctt gac gcg cac ctg tgc aag cac acg 370
Asn Tyr Ser Lys Ala Trp Lys Leu Asp Ala His Leu Cys Lys His Thr
          50                      55                      60                      65

ggg gag aga cca ttt gtt tgt gac tat gaa ggg tgt ggc aag gcc ttc 418
Gly Glu Arg Pro Phe Val Cys Asp Tyr Glu Gly Cys Gly Lys Ala Phe
          70                      75                      80

atc agg gac tac cat ctg agc cgc cac att ctg act cac aca gga gaa 466
Ile Arg Asp Tyr His Leu Ser Arg His Ile Leu Thr His Thr Gly Glu

```

9/14

feuille rectifiée

85	90	95	
aag ccg ttt gtt tgt gca gcc act ggc tgt gat caa aaa ttc aac aca Lys Pro Phe Val Cys Ala Ala Thr Gly Cys Asp Gln Lys Phe Asn Thr 100 105 110			514
aaa tca aac ttg aag aaa cat ttt gaa cgc aaa cat gaa aat caa caa Lys Ser Asn Leu Lys Lys His Phe Glu Arg Lys His Glu Asn Gln Gln 115 120 125			562
aaa caa tat ata tgc agt ttt gaa gac tgt aag aag acc ttt aag aaa Lys Gln Tyr Ile Cys Ser Phe Glu Asp Cys Lys Lys Thr Phe Lys Lys 130 135 140 145			610
cat cag cag ctg aaa atc cat cag tgc cag cat acc aat gaa cct cta His Gln Gln Leu Lys Ile His Gln Cys Gln His Thr Asn Glu Pro Leu 150 155 160			658
ttc aag tgt acc cag gaa gga tgt ggg aaa cac ttt gca tca ccc agc Phe Lys Cys Thr Gln Glu Gly Cys Gly Lys His Phe Ala Ser Pro Ser 165 170 175			706
aag ctg aaa cga cat gcc aag gcc cac gag ggc tat gta tgt caa aaa Lys Leu Lys Arg His Ala Lys Ala His Glu Gly Tyr Val Cys Gln Lys 180 185 190			754
gga tgt tcc ttt gtg gca aaa aca tgg acg gaa ctt ctg aaa cat gtg Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His Val 195 200 205			802
aga gaa acc cat aaa gag gaa ata cta tgt gaa gta tgc cgg aaa aca Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys Thr 210 215 220 225			850
ttt aaa cgc aaa gat tac ctt aag caa cac atg aaa act cat gcc cca Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala Pro 230 235 240			898
gaa agg gat gta tgt cgc tgt cca aga gaa ggc tgt gga aga acc tat Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr Tyr 245 250 255			946
act act gtg ttt aat ctc caa agc cat atc ctc tcc ttc cat gag gaa Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu Glu 260 265 270			994
agc cgc cct ttt gtg tgt gaa cat gct ggc tgt ggc aaa aca ttt gca Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe Ala 275 280 285			1042
atg aaa caa agt ctc act agg cat gct gtt gta cat gat cct gac aag Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp Lys 290 295 300 305			1090
aag aaa atg aag ctc aaa gtc aaa aaa tct cgt gaa aaa cgg agt ttg Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser Leu 310 315 320			1138

gcc tct cat ctc agt gga tat atc cct ccc aaa agg aaa caa ggg caa 1186  
 Ala Ser His Leu Ser Gly Tyr Ile Pro Pro Lys Arg Lys Gln Gly Gln  
                   325                                  330                                  335

ggc tta tct ttg tgt caa aac gga gag tca ccc aac tgt gtg gaa gac 1234  
 Gly Leu Ser Leu Cys Gln Asn Gly Glu Ser Pro Asn Cys Val Glu Asp  
                   340                                  345                                  350

aag atg ctc tcg aca gtt gca gta ctt acc ctt ggc taa 1273  
 Lys Met Leu Ser Thr Val Ala Val Leu Thr Leu Gly  
                   355                                  360                                  365

<210> 2  
 <211> 365  
 <212> PRT  
 <213> Human

<400> 2  
 Met Asp Pro Pro Ala Val Val Ala Glu Ser Val Ser Ser Leu Thr Ile  
           1                                  5                                  10                                  15

Ala Asp Ala Phe Ile Ala Ala Gly Glu Ser Ser Ala Pro Thr Pro Pro  
                   20                                  25                                  30

Arg Pro Ala Leu Pro Arg Arg Phe Ile Cys Ser Phe Pro Asp Cys Ser  
           35                                  40                                  45

Ala Asn Tyr Ser Lys Ala Trp Lys Leu Asp Ala His Leu Cys Lys His  
           50                                  55                                  60

Thr Gly Glu Arg Pro Phe Val Cys Asp Tyr Glu Gly Cys Gly Lys Ala  
           65                                  70                                  75                                  80

Phe Ile Arg Asp Tyr His Leu Ser Arg His Ile Leu Thr His Thr Gly  
                   85                                  90                                  95

Glu Lys Pro Phe Val Cys Ala Ala Thr Gly Cys Asp Gln Lys Phe Asn  
                   100                                  105                                  110

Thr Lys Ser Asn Leu Lys Lys His Phe Glu Arg Lys His Glu Asn Gln  
           115                                  120                                  125

Gln Lys Gln Tyr Ile Cys Ser Phe Glu Asp Cys Lys Lys Thr Phe Lys  
           130                                  135                                  140

Lys His Gln Gln Leu Lys Ile His Gln Cys Gln His Thr Asn Glu Pro  
           145                                  150                                  155                                  160

Leu Phe Lys Cys Thr Gln Glu Gly Cys Gly Lys His Phe Ala Ser Pro  
                   165                                  170                                  175

Ser Lys Leu Lys Arg His Ala Lys Ala His Glu Gly Tyr Val Cys Gln  
                   180                                  185                                  190

Lys Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His

195	200	205
Val Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys		
210	215	220
Thr Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala		
225	230	235
Pro Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr		
245	250	255
Tyr Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu		
260	265	270
Glu Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe		
275	280	285
Ala Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp		
290	295	300
Lys Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser		
305	310	315
Leu Ala Ser His Leu Ser Gly Tyr Ile Pro Pro Lys Arg Lys Gln Gly		
325	330	335
Gln Gly Leu Ser Leu Cys Gln Asn Gly Glu Ser Pro Asn Cys Val Glu		
340	345	350
Asp Lys Met Leu Ser Thr Val Ala Val Leu Thr Leu Gly		
355	360	365

&lt;210&gt; 3

&lt;211&gt; 1273

&lt;212&gt; ADN

&lt;213&gt; Human

&lt;400&gt; 3

```

atgcgcagca gggcgccga cgcggggcgg tgcttggtga ccgcgcgcgc tcccgaagt 60
gtgccggcgt cgcgcgaagg ttcagcaggg agccgtgggc cggcgcgcc ggttcccggc 120
acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccttgga 180
tccgccggcc gtggtcgccg agtcggtgtc gtccttgacc atcgccgacg cgttcattgc 240
agccggcgag agtcagctc cgaccccgcc gcgccccgcg cttcccagga ggttcattcg 300
ctccttcctt gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 360
caagcacacg ggggagagac catttgtttg tgactatgaa ggggtgtggca aggccttcac 420
cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttgtttg 480
tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aacttgaaga aacattttga 540

```

acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 600  
 ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 660  
 caagtgtacc caggaaggat gtgggaaaca ctttgcata cccagcaagc tgaaacgaca 720  
 tgccaaggcc cagcagggct atgtatgtca aaaaggatgt tcctttgtgg caaaaacatg 780  
 gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaataactat gtgaagtatg 840  
 ccggaanaaca tttaaacgca aagattacct taagcaacac atgaaaactc atgccccaga 900  
 aagggatgta tgctcgtgtc caagagaagg ctgtggaaga acctatacta ctgtgtttaa 960  
 tctccaaagc catatcctct ccttccatga ggaaagccgc ccttttgtgt gtgaacatgc 1020  
 tggctgtggc aaacatttg caatgaaaca aagtctcact aggcattgtg ttgtacatga 1080  
 tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1140  
 ctctcatctc agtggatata tccctcccaa aaggaaacaa gggcaaggct tatctttgtg 1200  
 tcaaaacgga gagtcacca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1260  
 tacccttggc taa 1273

<210> 4  
 <211> 1213  
 <212> ADN  
 <213> Human

<400> 4  
 gtgccggcgc cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccgcc 60  
 acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccttggc 120  
 tccgccggcc gtggtegccg agtcgggtgtc gtccttgacc atcgccgacg cgttcattgc 180  
 agccggcgag agctcagctc cgaccccgcc gcgccccgcg ctccccagga ggttcatctg 240  
 ctcttccct gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 300  
 caagcacacg ggggagagac catttgtttg tgactatgaa ggggtgtggc aggccttcat 360  
 cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttgtttg 420  
 tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aacttgaaga aacattttga 480  
 acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 540  
 ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 600  
 caagtgtacc caggaaggat gtgggaaaca ctttgcata cccagcaagc tgaaacgaca 660  
 tgccaaggcc cagcagggct atgtatgtca aaaaggatgt tcctttgtgg caaaaacatg 720

gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaatactat gtgaagtatg 780  
 ccggaaaaca tttaaagca aagattacct taagcaacac atgaaaactc atgccccaga 840  
 aagggatgta tgtcgtgtc caagagaagg ctgtggaaga acctatacta ctgtgtttaa 900  
 tctccaaagc catatcctct ccttccatga ggaaagccgc cttttgtgt gtgaacatgc 960  
 tggctgtggc aaaacatttg caatgaaaca aagtctcact aggcattgtg ttgtacatga 1020  
 tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1080  
 ctctcatctc agtggatata tccctcccaa aaggaaacaa gggcaaggct tatctttgtg 1140  
 tcaaaacgga gagtcacca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1200  
 tacccttggc taa 1213

<210> 5  
 <211> 34  
 <212> ADN  
 <213> Human

<400> 5  
 cggggtacca aaaatgcgca gcagcggcgc cgac

34

<210> 6  
 <211> 21  
 <212> ADN  
 <213> Human

<400> 6  
 tccttccctg actgcagcgc c

21

<210> 7  
 <211> 20  
 <212> ADN  
 <213> Human

<400> 7  
 tgcacagggtg cgcgtcaagc

20

<210> 8  
 <211> 20  
 <212> ADN  
 <213> Human

<400> 8  
 cacaaacaaa tggctctctcc

20

<210> 9  
 <211> 30



14/14

<212> ADN  
<213> Human

<400> 9  
cggtctagat tagccaaggg taagtactgc

30

<210> 10  
<211> 30  
<212> ADN  
<213> Human

<400> 10  
cctcccgggg ccaagggtaa gtactgcaac

30

**THIS PAGE BLANK (USPTO)**